

AD-A088 221

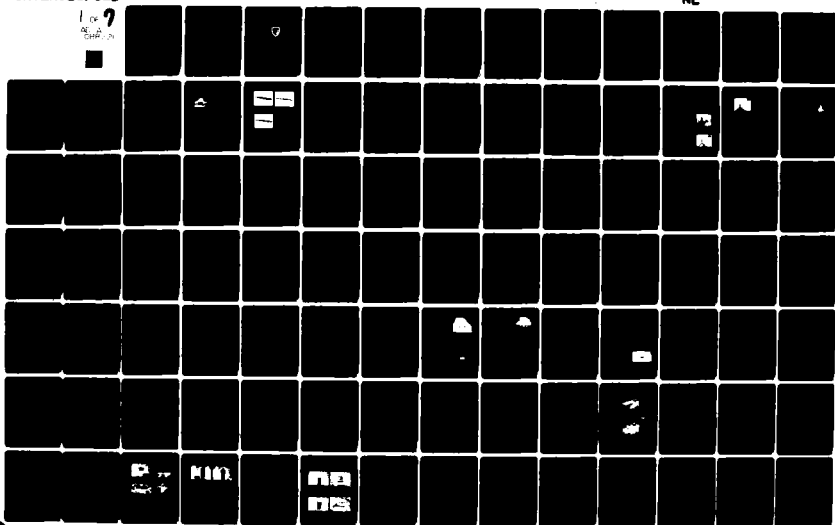
ARMY ELECTRONICS COMMAND FORT MONMOUTH NJ
PROCEEDINGS OF THE ANNUAL FREQUENCY CONTROL SYMPOSIUM (31ST) HE-ETC(U)
1977

F/G 9/5

UNCLASSIFIED

NL

1 of 9
SEARCHED
SERIALIZED



LEVEL III

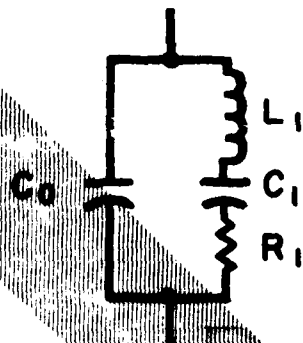
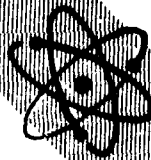
①

PROCEEDINGS
OF THE

3074 A046089

31st ANNUAL SYMPOSIUM ON FREQUENCY CONTROL
1977

AD A088221



DISTRIBUTION STATEMENT A
Approved for public release;
Distribution unlimited

1-3 June 1977

A

U.S. Army Electronics Command
Fort Monmouth, New Jersey

DDC FILE COPY



August 6, 1980

Mr. John Ashley
National Technical Information Service
Sills Building
5285 Port Royal Road
Springfield, VA 22161

Dear Mr. Ashley:


As the copyright holder for the Proceedings of the Annual Symposium on Frequency Control, which is sponsored by the U.S. Army Electronics Command, this letter on behalf of EIA is to authorize the NTIS to reproduce and distribute at its normal price schedule copies of the following Proceedings:

24th Edition - 1970 - AD 746210
31st Edition - 1977

I understand an AD number has already been established for the 24th Edition (as above) and I am enclosing a copy of the proceeding of the 31st for your convenience.

Please supply price and AD number information as soon as possible for inclusion in our 34th Edition which is soon to be released.

Sincerely,


Herbert J. Rowe
Vice President

cc: Dr. Erich Hafner (USAECOM)

file

(6) PROCEEDINGS

of the

~~THIRTY-FIRST~~ ANNUAL FREQUENCY CONTROL SYMPOSIUM (31st)
Held at ¹⁹⁷⁷
Atlantic City, New Jersey on 1-3
Sponsored by
June 1977.



11/1977

12/633

U.S. ARMY ELECTRONICS COMMAND
Fort Monmouth, N.J.
Major General John K. Stoner, Jr.
Commanding

ELECTRONICS TECHNOLOGY AND DEVICES LABORATORY
Dr. C. G. Thornton
Director

© Electronic Industries Association 1977
All rights reserved
Printed in U.S.A.

Copies of the Proceedings are available from:

Electronic Industries Association
2001 Eye Street, N.W.
Washington, D.C. 20006

PRICE PER COPY: \$10.00

AUG 20 1980
A

634P

None of the papers contained in the Proceedings may be reproduced in whole or in part, except for the customary brief abstract, without permission of the author, and without due credit to the Symposium.

1 - 3 June 1977
Howard Johnson's Regency Hotel
Atlantic City, New Jersey

DESTROYED NON SERIALIZED
Approved for Release

037620

THIRTY-FIRST ANNUAL FREQUENCY CONTROL SYMPOSIUM

Sponsored By

U.S. ARMY ELECTRONICS COMMAND
ELECTRONICS TECHNOLOGY AND DEVICES LABORATORY
Fort Monmouth, New Jersey

1 - 3 June 1977

Howard Johnson's Regency Hotel
Atlantic City, New Jersey

*Letter on file
Copyright*

SYMPOSIUM EXECUTIVE COMMITTEE

Chairman
Executive Assistant
Executive Secretary

Dr. Erich Hafner
Dr. John Vig
Mrs. Lee Hildebrandt

A

TECHNICAL PROGRAM COMMITTEE

Dr. Erich Hafner-Chairman
U.S. Army Electronics Command

Mr. Edward Alexander
Bell Laboratories

Mr. Jack Kusters
Hewlett-Packard

Dr. Art Ballato
U.S. Army Electronics Command

Mr. Ted Lukaszek
U.S. Army Electronics Command

Mr. Andy Chi
NASA - Goddard Space Flight Center

Mr. Pete Maresca
U.S. Army Satellite Comm. Agency

Dr. Lewis Claiborne
Texas Instruments

Mr. Dennis Reifel
Motorola

Dr. Cecil Costain
National Research Council

Mr. Lauren Rueger
APL, Johns Hopkins University

Mr. Victor Friedrich
Ofc, Asst. Secy. of the Army (R&D)

Mr. Stan Schodowski
U.S. Army Electronics Command

Dr. Helmut Hellwig
National Bureau of Standards

Mr. John Sherman
General Electric

Mr. John Holmbeck
Northern Engineering Laboratories

Mr. Bob Stone
Naval Research Laboratory

Mr. Alfred Kahan
RADC, Air Force

Mr. Ted Viars
U.S. Army Electronics Command

Mr. Eric Kentley
C. R. Snelgrove

Dr. John Vig
U.S. Army Electronics Command

Dr. Jim King
Sandia Laboratories

Dr. Gernot Winkler
U.S. Naval Observatory

CHAIRMEN FOR TECHNICAL SESSIONS

OPENING REMARKS

Mr. Victor L. Friedrich, Office of the Assistant Secretary of the
Army (R&D)

FORCE AND TEMPERATURE SENSITIVITY OF QUARTZ CRYSTAL RESONATORS

Dr. R. N. Thurston, Bell Laboratories

QUARTZ CRYSTAL RESONATORS DESIGN, MEASUREMENT AND APPLICATION

Mr. Dennis Reifel, Motorola

RESONATOR PROCESSING

Dr. Martin Bloch, Frequency Electronics

FUNDAMENTAL PROPERTIES OF MATERIALS FILTERS

Mr. Edward J. Alexander, Bell Telephone Laboratories

SURFACE ACOUSTIC WAVE RESONATORS AND FILTERS

Dr. Lewis Claiborne, Texas Instruments

OSCILLATOR SPECIFICATION AND MEASUREMENT

Dr. Helmut Hellwig, National Bureau of Standards

OSCILLATORS

Mr. Marvin E. Frerking, Collins Radio

FREQUENCY AND TIME CONTROL ASPECTS OF DIGITAL COMMUNICATION

Dr. Cecil Costain, National Research Council

TIME DISTRIBUTION AND SYNCHRONIZATION ATOMIC AND MOLECULAR FREQUENCY STANDARDS

Mr. Roger Easton, Naval Research Laboratory

FAR INFRARED/SUBMILLIMETER WAVE DEVICES FOR FREQUENCY STANDARDS - NOVEL RESONATORS

Dr. Len Cutler, Hewlett-Packard

*Reports are filed under
the following subheadings:*

TABLE OF CONTENTS

	Page No.
<u>Opening Remarks</u>	
- Victor L. Friedrich, Office of Assistant Secretary of the Army (R&D)	1
<u>Force and Temperature Sensitivity of Quartz Crystal Resonators</u>	
TTC's - Further Developmental Results	
- J. A. Kusters, C.A. Adams and H. Yoshida, Hewlett-Packard Laboratories, and J. G. Leach, Hewlett-Packard Company.....	3
The Force-Frequency Effect in Doubly Rotated Quartz Resonators	
- A. Ballato and T. Lukaszek, US Army Electronics Technology & Devices Laboratory (ECOM) and E.P. EerNisse, Sandia Laboratories.....	8
Amplitude - Frequency Behavior of Doubly Rotated Quartz Resonators	
- J. J. Gagnepain, C.N.R.S., J.C. Poncot, E.N.S.C.M.B., and C. Pegeot, Quartz de Electronique.....	17
Temperature Induced Frequency Changes in Electroded AT-Cut Quartz Thickness-Shear Resonators	
- H. F. Tiersten and B. K. Sinha, Rensselaer Polytechnic Institute.....	23
The Influence of Support Configuration on the Acceleration Sensitivity of Quartz Resonator Plates	
- P.C.Y. Lee and Kuang-Ming Wu, Princeton University.....	29
<u>Quartz Crystal Resonators Design, Measurement and Application</u>	
X1 and X3 Flexure, Face-Shear, Extension, Thickness-Shear, and Thickness- Twist Modes in Rectangular Rotated Y-Cut Quartz Plates	
- T. R. Meeker, Bell Telephone Laboratories.....	35
An Analysis of Overtone Modes in Contoured Crystal Resonators	
- H. F. Tiersten, Rensselaer Polytechnic Institute, and R. C. Smythe, Piezo Technology Inc.....	44
4 MHZ AT-Cut Strip Resonator for Wrist Watch	
- M. Onoe, University of Tokyo, and K. Kamada, M. Okazaki, F. Tajika and N. Manabe, Nihon Dempa Kogyo Co., Ltd.....	48
DT-Cut Torsional Resonators	
- J. Hermann, Centre Electronique Horloger S.A.....	55
Quartz Crystal Accelerometer Insensitive to Temperature Variation	
- M. Onoe, University of Tokyo, and K. Furusawa, Meisei Electric Co., Ltd., and S. Ishigami, T. Sase and M. Sato, Tokyo Denpa Co., Ltd.....	62
Using the X-Y Flexure Watch Crystal as a Pressure-Force Transducer	
- A. Genis and D. E. Newell, Northern Illinois University.....	71
Methods for Production Screening for Anomalous Responses in Quartz Crystals Intended for High Reliability Applications	
- P. F. Godwin, Jr. and G. L. Snider, TRW Systems.....	78
Extending the Frequency Range of the Transmission Line Method for the Measurement of Quartz Crystals Up to 250 MHz	
- R. Fischer and L. Schulzke, KVG - Germany.....	96
Ovenless Activity Dip Tester	
- A. Ballato and R. Tilton, US Army Electronics Technology and Devices Laboratory (ECOM).....	102

	<u>Page No.</u>
Measurement of the Characteristic Frequency of an AT-Cut Plate - J. H. Sherman, Jr., General Electric Company.....	108
A New Quality Evaluation Method of Raw Quartz by Mesasuring the Q-Value of Y-Bar Resonator - H. Fukuyo, N. Oura and F. Shishido, Tokyo Institute of Technology.....	117
→ <u>Resonator Processing</u> ;	
Accoustic Bulk Wave Resonators and Filters Operating in the Fundamental Mode at Frequencies Greater than 100 MHZ - M. Berte, Thompson - CFS	122
The Relationship Between Quartz Surface Morphology and the Q of High Frequency Resonators - R. N. Castellano, T.R. Meeker and R.C. Sundahl, Bell Telephone Laboratories.....	126
Chemically Polished Quartz - J. R. Vig, J. W. LeBus and R. L. Filler, US Army Electronics Command.....	131
Aging Analysis of Quartz Crystal Units with Ti Pd Au Electrodes - G. L. Dybwad, Bell Telephone Laboratories.....	144
A New "Electrodeless" Resonator Design - R. J. Besson, E.N.S.C.M.B.....	147
Making Doubly Rotated Quartz Plates - W. L. Bond, Stanford University, and J. A. Kusters, Hewlett-Packard.....	153
<u>Fundamental Properties of Materials</u>	
The Electroelastic Tensor and Other Second-Order Phenomena in Quasilinear Interpretation of the Polarizing Effect with Thickness Vibrations of a-Quartz Plates - C. K. Hruska, York University.....	159
Coercive Stress for Ferrobielastic Twinning in Quartz - T. L. Anderson, R. E. Newnham and L. E. Cross, Pennsylvania State University.....	171
Hydrothermal Synthesis of Aluminum Metaphosphate - E.D. Kolb and R. A. Laudise, Bell Laboratories.....	178
Temperature Compensated Cuts of Berlinite and β -Eucryptite for SAW Devices - 1Lt. R. M. O'Connell and P. H. Carr, Hanscom AFB.....	182
<u>Filters</u> •	
LCD Recursive Filter for MTI Applications - W. L. Eversole, W. H. Bailey and P. L. Ham, Texas Instruments.....	187
Unwanted Modes in Monolithic Crystal Filters - G. T. Pearman and R.C. Rennick, Bell Telephone Laboratories.....	191
Multi-Mode Stacked Crystal Filter - C. M. Stearns, S. Wanuga, S. W. Tehon, General Electric, and A. Kachelmyer, University of Minnesota.....	197
128 kHz Pole-Type Mechanical Channel Filter - K. Yakuwa, S. Okuda, K. Shirai and Y. Kasai, Fujitsu Limited.....	207
Lithium Tantalate Channel Filters for Multiplex Telephony - T. Arranz, C.E.P.E.....	213

Surface Acoustic Wave Resonators and Filters

	Page No.
1.2 GHz Temperature-Stable SAW Oscillator - R.D. Weglein, Hughes Research Laboratories.....	225
Tuning Quartz SAW Resonators by Opening Shorted Reflectors - R. C. Rosenfeld, T. F. O'Shea, and S. H. Arneson, Motorola, Inc.....	231
Tunable Variable Bandwidth/Frequency SAW Resonators - J. S. Schoenwald, Teledyne MEC.....	240
Deeply Etched SAW Resonators - C. A. Adams and J. A. Kusters, Hewlett-Packard.....	246
Piezoelectric Shear Surface Wave Grating Resonators - B. A. Auld and Bing-Hui Yeh, Stanford University.....	251
A New Class of Quartz Crystal Oscillator Controlled by Surface-Skimming Bulk Waves - T. I. Browning and M. F. Lewis, RSRE.....	258
Interdigital Transducers--A Means of Efficient Bulk Wave Excitation - K. H. Yen, K. L. Wang, R. S. Kagiwada and K. F. Lau, TRW Defense and Space Systems Group.....	266
Surface Acoustic Wave Scattering From a Groove in Y-Z LiNbO ₃ - S. D. Wu and H. S. Tuan, State University of New York.....	271
UHF Range SAW Filters Using Group-Type Uni-Directional Interdigital Transducers - J. Otomo, S. Nishiyama, Y. Konno, Nihon Dempa Kogyo Co., Ltd., and S. Shibayama, Tohoku University.....	275
Progress Report on Surface Acoustic Wave Device MMT - II - A. R. Janus and L. Dyal III, Hughes Aircraft Company.....	281
The Versatility of the "In-Line" SAW Chirp Filter - W. J. Skudera, Jr., US Army Electronics Command.....	285
<u>Oscillator Specification and Measurement</u>	
Oscillator Specifications: A Review of Classical and New Ideas - J. Rutman, L.P.T.F.....	291
Specification and Measurement of Oscillator Phase Noise Instability - W. C. Lindsey, University of Southern California, and C. M. Chie, LinCom Corporation.....	302
Estimation of the Two-Sample Variance with a Limited Number of Data - P. Lesage and C. Audoin, C.N.R.S.....	311
Prediction Error Analysis of Atomic Frequency Standards - D. B. Percival, US Naval Observatory.....	319
Models and Predictions for the Realization of Time Scales - U. Hubner, Physikalisch-Technische Bundesanstalt.....	327
Accurate Measurements of Spectral Density of Phase Noise in Devices - F. L. Walls and S. R. Stein, National Bureau of Standards.....	335
Frequency and Time Domain Stability of the Rb ⁸⁷ Maser and Related Oscillators: A Progress Report - J. Vanier, M. Tetu and R. Brousseau, Universite Laval.....	344

	<u>Page No.</u>
Automating Phase Noise Measurements in the Frequency Domain - A. L. Lance, W. D. Seal, F. G. Mendoza and N. W. Hudson, TRW Defense and Space Systems Group.....	347
<u>Oscillators</u>	
Current Developments in SAW Oscillator Stability - T. E. Parker, Raytheon Research Division.....	359
SAW Oscillators for Phase Locked Applications - T. R. Joseph, Hughes Aircraft Company.....	365
300 MHz Oscillators Using SAW Resonators and Delay Lines - E. J. Staples and T. C. Lim, Rockwell International.....	371
Ultra-Stable LC Oscillators and Their Applications in Metrology - C. T. Van Degrift, National Bureau of Standards.....	375
Frequency Synthesizers for Airborne Radars - F. W. Hopewood, J. P. Muhlbaier and H. Rossman, Westinghouse Electric Corporation.....	385
Frequency Tolerance Limitations with Logic Gate Clock Oscillators - J. D. Holmbeck, Northern Engineering Laboratories, Inc.....	390
The MXO - Monolithic Crystal Oscillator - T. Luxmore and D. E. Newell, CTS Knights, Inc.....	396
Linear Crystal Controlled FM Source for Mobile Radio Application - R. Arakelian, Iran Electronics Industries, and M. M. Driscoll, Westinghouse Electric Corporation.....	400
Design Considerations for a Digitally Temperature Compensated Crystal Oscillator - P. J. Scott, Motorola Communications Group.....	407
Precision Oscillators Flown on the LES-8/9 Spacecraft - H. S. Babbitt, III, Massachusetts Institute of Technology.....	412
A Fast Warmup Oscillator for the GPS Receiver - J. Ho and M. Bloch, Frequency Electronics, Inc.....	421
<u>Frequency and the Control Aspects of Digital Communication</u>	
Synchronization Methods for Frequency - And Time-Division- Multiplex Networks - H. L. Hartman, Technische Universitat Braunschweig.....	429
Frequency Control and Digital Network Synchronization - M. I. Spellman, J. B. Cain and D. B. Bradley, Harris Electronic Systems Division.....	436
Frequency Coordination in a Network of Integrated Switches - H. A. Sunkenberg and Dr. M. J. Ross, GTE Sylvania Incorporated.....	448
Results of Investigations for the Clock Frequency Control and Distribution System in the Digital Telephony and Data Networks of the Deutsche Bundespost and Future Plans - W. R. Slabon, Fernmeldetechnisches Zentralamt der Deutschen Bundespost.....	455

	Page No.
Synchronization of a Digital Network Using a Position Lock and HMS System, - P. K. Verma and C. Lemieux, Bell Canada, and E. Daams, National Research Council.....	463
An Overview of TDMA for Digital Satellite Communications and the Censar Synchronization Experiment - K. E. Brown and P. P. Nuspl, Canadian Department of Communication.....	465
Frequency Control and Timing Requirements for Communications Systems - P. Kartaschoff, Swiss Post Office.....	478
Generation of Base-Band Frequencies for FDM and TDM Telecommunications - E. P. Graf and B. Walther, Oscilloquartz S.A.....	484
<u>Time Distribution and Synchronization</u>	
Time Determination for Spacecraft Users of the Navstar Global Positioning System (GPS) - T. J. Grenchik, NASA/GSFC, and B. T. Fang, EG&G.....	489
Transcontinental and Intercontinental Portable Clock Time Comparison - H. Hellwig, D. W. Allan, S. R. Stein and K. A. Prichard, National Bureau of Standards.....	495
Distant Comparison of Stable Frequency Standards by Means of the Transmission of a Beat Note Between the Carrier of a TV Broadcast Signal and a Frequency Synthesized from the Frequency Standards - A. Gabry, G. Faucheron and B. Dubouis, C.N.E.T., and P. Petit, LHA Universite Paris-SUD.....	499
Study of L.F. and V.L.F. Time Signals by Digital Method - F. Guillaume, J. C. Lievin and J. De Prins, Universite Libre de Bruxelles.....	503
<u>Atomic and Molecular Frequency Standards</u>	
Design of a Spacecraft Hydrogen Maser - E. S. Sabisky, RCA, and H. E. Peters, Consultant.....	510
Hydrogen Maser with a Double Configuration Bulb for Wall Shift Measurements in the Temperature Range 25-120°C - P. Petit, M. Desaintfuscien and C. Audoin, CNRS.....	520
A New Generation of SAO Hydrogen Masers - M. W. Levine, R. Vessot, E. Mattison, G. Nystrom, T. Hoffman and E. Blomberg, Smithsonian Astrophysical Observatory.....	525
Zeeman Effects on H and Rb Masers - G. Busca, J-Y Savard, S. Rovea and J. Vanier, Universite Laval, and M. Desaintfuscien, P. Petit and C. Audoin, Universite de Paris.....	535
Development of a Cesium Beam Clock for Satellite Application - J. George and A. I. Vulcan, Frequency Electronics, Inc.....	542
New Cesium Beam Frequency Standards for Flight and Ground Applications - T. K. Gregory, Frequency and Time Systems, Inc.....	551
Some Causes and Cures of Frequency Instabilities (Drift & Noise) in Cesium Beam Frequency Standards - D. W. Allan, H. Hellwig, S. Jarvis, Jr., D. A. Howe and R. M. Garvey, National Bureau of Standards.....	555

	<u>Page No.</u>
Results with the Special-Purpose Ammonia Frequency Standard - D. J. Wineland, D. A. Howe, and M. B. Mohler, National Bureau of Standards.....	562
<u>Far Infrared/Submillimeter Wave Devices for Frequency Standards - Novel Resonators .</u> ←	
Detecting and Mixing at FIR/Submillimeter Wavelengths With Submicron Size Schottky Barrier Diodes - M. McColl, D. T. Hodges, A. B. Chase and W. A. Garber, The Aerospace Corporation.....	574
A Narrow Output Linewidth Multiplier Chain for Precision Frequency Measurements in the 1 THz Region - E. Bava, A. De Marchi and A. Godone, Istituto Elettrotecnico Nazionale.....	578
A Study of Point-Contact Josephson Junctions for Use in Frequency Synthesis - A. S. Risley, National Bureau of Standards.....	583
Optical Electronics, Extension of Microwave Techniques Into The Optical Region - C. F. Davis, Jr., G. Elchinger, A. Sanchez, K. C. Liu and A. Javan, Massachusetts Institute of Technology.....	590
Progress in CO ₂ Laser Stabilization - C. Freed, MIT, Lincoln Laboratory.....	592
Electronic Tuning and Phase-Lock Techniques for Optically Pumped Far Infrared Lasers - S. R. Stein, National Bureau of Standards, and H. Van de Stadt, National Bureau of Standards and University of Colorado.....	601
Frequency Modulation of a Far Infrared CH ₃ F Laser by Stark Effect - R. Benedetti, A. Di Lieto, M. Inguscio, P. Minguzzi, F. Strumia and M. Tonelli, Istituto di Fisica dell 'Universita - Pisa.....	605
The Trapped Mercury Ion Frequency Standard - M. D. McGuire, NASA/GSFC.....	612
Clocks Based upon High Mechanical Q Single Crystals - D. F. McGuigan and D. H. Douglass, University of Rochester.....	616
INDEX OF AUTHORS.....	621
SPECIFICATIONS AND STANDARDS GERMANE TO FREQUENCY CONTROL.....	623
PROCEEDINGS ORDER INFORMATION.....	Inside Back Cover

Opening Remarks

Victor L. Friedrich
Deputy for Communications and Target Acquisition
Office of Assistant Secretary of the Army (R&D)
Washington, D.C.

With this symposium having completed a 30 year run and entering the 4th decade, I would like to commend the Electronics Technology and Devices Laboratory of the Electronics Command at Fort Monmouth for these very excellent series of annual symposia which they have so effectively sponsored. Having been a member of the Research Office of the Electronics Command when the symposia were started, I had the opportunity to participate in the meetings from their inception. As many of you know, the meetings were held in the Fort Monmouth area until the facilities became inadequate due to the sizeable participation in the meetings. The discussions at the early meetings centered largely on the various aspects of quartz, including crystal cuts, aging, of electrode loading, and, of course, the field of synthetic quartz growth was then in its early phase, and the papers and discussions dealt with experimental techniques for the growth of quality synthetic quartz in laboratory autoclaves.

Several years ago I had occasion to visit the Western Electric plant in Andover, Mass., where I toured their synthetic quartz facility. The sizeable underground autoclave installation and the racks of synthetically grown quartz crystals were evidence not only of the products of the researchers but also of the effectiveness of the information interchange which has been accomplished so extensively at these symposia.

In reviewing this year's program one finds that quartz still occupies the attention of the researchers, but its applications have expanded from frequency control and filters to many others, including the very active field of surface acoustic wave technology. This is an area in which many diverse applications are emerging, not the least of which are the advances which it is creating in signal processing, and signal processing in the communications and the radar field is very promising.

In the time that I have, I would like to dwell on some aspects of military communications electronics development of today and which in part are dependent on the area of frequency control.

If we examine the history of net type radio development, I am referring now to the VHF/FM manpack and vehicular type radios, we need but examine the radio frequency channel spacing to realize the effect that advances in technology have fostered. The early radios, those of the nineteen forties, used 100 kHz channels for FM communications. The next generation radios, those of the late fifties and still largely in use today, employ 50 kHz channels. The improvement in frequency stability technology coupled with advances in basic radio technology thus provided us with twice the radio communication capacity over the earlier equipment. We are now proceeding with development of the next generation radio system. Once again the radio frequency channel bandwidth is being halved, to 25 kHz. These radios will transmit and receive conventional analogue FM voice communications and will also communicate both digitized voice and data. Current frequency

control and frequency synthesis technology is capable of providing the stability and frequency tuning needs for the 25 kHz channel spacing. However, as we look further into the future, we find techniques which promise to further and substantially reduce the radio frequency channel bandwidth needed for effective and high quality voice communications. The application of sophisticated coding and signal processing techniques promises these further substantial reductions. As these techniques reach maturity and if hopefully they become cost effective, a corresponding need for improved frequency standards for tactical systems will become a requirement.

While decreased bandwidth is a desirable goal, there are several other equally important goals. Communications under severe signal interference is a challenging goal. Techniques such as rapid frequency hopping, that is, changing frequency many times each second, by sending short bursts of RF with each burst being transmitted at a different frequency on a pseudo random basis, or the use of spread spectrum both of which require a form of synchronization are potential solutions. These same techniques also are promising, at least to a limited extent, to fulfill a second need, that of concealed emissions or emissions with a low probability of intercept.

In the area of data transmission, as you know, there is appreciable commercial exploitation and application. In the military narrow band data is now being transmitted quite extensively and the need for transmission of broad band data is increasing. Methods for more efficient and effective transmission of distributed type multi-channel information are being explored with time division multiple access being given appreciable attention. In the distributed TDMA concept being explored for military use all users transmit on the same frequency during their time slot, which is, of course, an inherent feature of TDMA systems, and thus the data can be directed to any desired participating users. This is of particular value when data needs to be distributed to many users or when a number of users need to share or interchange data. Since the time slots throughout such a distributed system must be controlled or time synchronized within reasonable tolerance, the system will also provide relative position data among the users, some or all of whom may be on the move. This offers several advantages since each user can determine his position with respect to his neighboring users or the entire user community positions can be displayed centrally.

Another approach being examined for multiple user transmission in military applications does not require rigid channelization of the time domain. In this form of communication the data to be transmitted is packaged into packets, each containing say up to several thousand bits. Each packet has a header which contains the address and the routing. The communications system for tactical military applications would consist of a distributed array of radio repeaters placed through the area of coverage so as to overcome line

of sight limitations. No pre-assigned time slots are used, but instead, before a packet is transmitted, the system samples to determine if the circuit is free. In this way many users with small quantity data transmission needs can have access to the system. As the system becomes heavily loaded, the time to complete an end to end transmission may increase. However, the system is not limited to a discrete number or to dedicated time slots and is thus able to accommodate many distributed subscribers.

In the area of satellite communications, existing military satellite systems dedicate channel capacity to pre-assigned users. Specific frequency slots are allocated to each user of the system. While this provides for high channel accessibility and availability, it may however, result in low duty factor per channel and restricts the number of users who have access to the satellite system.

Recent developments in satellite communications technology will permit more efficient assignment of satellite capacity through a concept called Demand Assigned Multiple Access. This concept will permit assignment of satellite communications capacity to users as required, and on demand by sharing the time channels with other users.

Looking at the higher frequencies we find that the millimeter and submillimeter portion of the spectrum has seen increased R&D activity in the past few years. Applications in communications including wide band volume type communications for satellite use and for short haul terrestrial use and also point to point single channel communications are being addressed. Also for unique applications, investigations are underway for the use of frequencies in which propagation attenuation provides a measure of communications privacy or protection from intercept. In addition, high precision radars and guidance systems are being explored at these short wavelengths which can provide highly accurate position and velocity data as well as imagery similar to that achieved with photography. The advanced applications that millimeter and submillimeter wavelengths promise are causing increased componentry developments at these wavelengths. No doubt the need for precision frequency standards will become more dire as further applications in these portions of the spectrum become reality.

These are some of the technology areas we have under investigation. The role of electronics is an ever expanding role in military systems and the conduct of military operations is becoming increasingly more dependent upon the applications of advanced electronic technology. There is hardly a system, weapon or supporting, land, sea, air or space, that does not depend upon electronics as a key element in its operation. The effective and efficient utilization of the spectrum for these applications is a role that has been effectively accomplished by the frequency standards community and we look forward to your continuation of these efforts to achieve even greater performance.

TTC'S - FURTHER DEVELOPMENTAL RESULTS

John A. Kusters
Charles A. Adams
Henry Yoshida
Hewlett-Packard Laboratories
Palo Alto, California

and

Jerry G. Leach
Hewlett-Packard Company
Santa Clara, California

Summary

Further developmental work has been performed to determine the suitability of the thermal transient compensated crystal cut (TTC) for use as a precision frequency standard.

Current investigation is concentrated on a 10 MHz, 3rd overtone design. Of interest are data on temperature characteristics, frequency-drive level dependence, '2-g' tipover performance, short term stability, long term aging and thermal transient performance. The superior temperature performance and drive level independence shown in one diopter contoured units implies that oscillator and oven design can be greatly simplified with no reduction in precision time base performance.

Introduction

The work reported here on the TTC was initially inspired by a theoretical prediction of an orientation in quartz that was free of thermal transient effects, the TS cut^{1,2}. After confirming experimentally the existence of this cut³, the name was changed to the TTC cut with the concurrence of Holland. Essentially, the same orientation was predicted for another crystal cut that was compensated for mounting stresses and electrode stresses in the plane of the crystal, the SC cut⁴. Further experimental data⁵ has been presented on the TTC cut which further confirms Holland's predictions, but still only partially confirms those of EerNisse.

General

The work reported here was confined to the development of a 10 MHz, 3rd overtone design which is suitable for high precision, oven controlled timebases. From previous work,⁵ the orientation, (yxw1) 21.93°/34.11°, was determined to be the optimum cut. All crystal units are plano-convex, 15 mm. in diameter, with 11.4 mm. electrode diameter and polished on both major surfaces and edge. X-ray orientation of the finished unit is better than 1' of arc. Contour is held to a tolerance of ± 0.05 diopters. After fabrication, the blanks are base-plated, then installed in an in-line system, brazed to a two-point mount, frequency plated, baked, backfilled to 4 torr of He at 375°C, and finally sealed into a coldweld enclosure without further exposure to atmosphere. Sealed and unsealed units are shown in Fig. 1.

The initial goal was to develop a crystal capable of rapid heating in an oven without significant frequency overshoot. Fig. 2 is a plot of data taken using a 5 MHz AT cut in a fast warm-up oven. The oven normally reaches its operating temperature in 300 seconds. Frequency transients induced by thermal gradients causes the AT to overshoot its desired frequency for substantial time after the oven reaches its operating temperature. Fig. 3 shows the same data taken with the same oven using a TTC cut. When the oven reaches its design temperature, the oscillator is within 1 part in 10^9 of its design frequency. Within several seconds, the oscillator is within the system measurement limits of ± 2 parts in 10^9 .

A further advantage of the TTC is shown in Fig. 4. This shows the temperature-frequency performance of three representative cuts, the AT, BT and TTC cut, all operating at the same turnover temperature of about 80°C. The TTC is within a part in 10^9 over a temperature range of $\pm 0.8^\circ\text{C}$. This superior temperature performance, coupled with its fast warmup capabilities, were predictable on the basis of the original theory¹⁻². The further results presented here were not expected when the study started.

Contour Dependence

TTC units were initially made in three different contours, one, two and three diopters. Table I shows the mean resistance and Q data as measured on these units.⁶ These measurements indicate that minimum resistance and maximum Q do not occur at the same contour as they did for the 5 MHz, 5th overtone units.³ Further testing produced the frequency-drive level dependency results shown in Fig. 5. One diopter units show virtual independence of the crystal drive current. For comparison, typical data for AT and BT cuts are also shown.

Since the TTC is a doubly-rotated cut, the acoustic mode spectra is expected to be quite messy. Fig. 6 shows the measured mode spectra for the three test contours. As expected, increasing the contour permits a greater separation between the desired response, and the beginning of the anharmonic responses. However, this also increases the frequency-drive level dependence and the series resistance. A good compromise seems to be the two diopter contour. This yields maximum Q with a moderate increase in resistance and drive level dependence. Further, the first anharmonic mode is almost 125 KHz above the desired mode, reducing possible problems of nonlinear coupling or activity dips.

Mounting Dependence

Of primary importance is the performance of the TTC cut with respect to the mounting location. The rotational angle convention used is shown in Fig. 7.

Table II presents data taken on various 2 diopter, plano-convex units as a function of the mounting location. As expected, the crystal parameters shown are at best a weak function of the mounting location. The data shown is the mean of the various units. The frequency-temperature coefficients are referenced at the turnover temperature.

Fig. 8 indicates the amount of thermal transient compensation achieved as a function of mount location. With this orientation, the average compensation is some 200 times better than the AT or the BT. As a result, the temperature scan necessary to obtain this data was about 25°C/minute. The data is not totally conclusive, but does indicate that the best compensation is achieved at a rotational orientation of about -15°. The data also shows quite strongly that even though the helium backfill in the can causes gaseous conduction to be the primary heat transfer mechanism, there is still a significant transfer of thermal energy through the mounting pins.

The mounting dependence for '2-g' tipover testing is shown in Fig. 9. In this and in Fig. 8, the solid curve shown is strictly a guess and is intended to be only representational, not an exact locus of data. The results observed to date indicate the best response to a 2-g change occur at rotational angle of about 20°. The actual response for a 15° mount is shown in Fig. 10. Preliminary squeeze testing indicates that the TTC does have a zero frequency-stress dependence at about 20° from the Z-Z' axis. The same zero dependence occurs at a point about 20° from the X-X' axis. At this orientation, however, the '2-g' dependence is almost at a maximum. All of this seems to conflict with the prediction of the SC cut that zero frequency-stress effects occur at ±45° from the Z-Z' axis, and that these two mounting axes should have identical behavior.

Unfortunately, the desired response for the thermal transient behavior and the '2-g' performance do not occur at the same rotational angle. Again, a compromise seems to be necessary. An alternative route, not yet taken, would be to redesign the mount and the thermal path through the mounting pins so that the two effects are minimized at the same mounting location. The compromise chosen was to restrict the remainder of the measurements to units mounted on the Z-Z' axis.

Stability and Aging

The short term stability, as expressed by the Allen Variance, was measured for a group of crystals. A typical result is shown in Fig. 11. Even though these crystals have lower Q than comparable AT and BT units, the short term stability is measurably better. Typical values for 1 second samples are: AT, about 5×10^{-12} ; BT, about 2.5 to 3×10^{-12} . Measured values for the TTC range from 1.24 to 1.87×10^{-12} . Measurements for sampling times less than 0.1 seconds have been limited by oscillator performance, and are not considered representative of the crystal.

Long term aging has also been determined for a limited group of crystals. The initial aging rate, after two days on the aging racks at a nominal turn-

over of 75°C, is 1×10^{-10} per day. This is fully comparable to current BT units made in the same manner. The BT units, ultimately after about 100 days, reach an aging rate of about 3×10^{-12} per day. Insufficient time has elapsed to indicate the final aging rate that the TTC's will reach, but at this time it is expected to be comparable to the BT units.

Conclusions

A generalized comparison between the TTC and similar conventional cuts, such as the AT and the BT, is shown in Fig. 12. We believe that the data presented in this paper and in its two predecessors^{3,5} shows that the TTC is superior in fast warm-up characteristics, in its frequency-temperature behavior, and in its frequency-drive level dependence. Further, the TTC is significantly better in its measured short term stability and in the observed '2-g' tipover tests. Finally, no device tested to-date has shown any significant activity dips. In most testing cycles, any frequency variation to a part in 10^9 , or impedance deviation of one-half ohm or greater caused by activity dips would have been observed. None were.

The data also implies that values related to the simple geometry of the quartz blank, such as shunt capacitance, remain essentially unchanged. Also, preliminary indications are that the long term aging of the TTC will be comparable to that observed on similar 10 MHz, 3rd overtone, BT's.

The data also shows that the TTC will yield lower Q values, higher series resistance values, and a richer, more complicated frequency spectrum. The most serious drawback to the full implementation of the TTC may be its tight manufacturing tolerances. The slopes of the lower turnover temperature, as a function of the orientation angles for the TTC are: -75.3°C/°φ and -248°C/°θ. As the TTC is not located on a convenient family of atomic planes, the amount of extra X-ray orientation required⁷ will probably restrict this cut to uses requiring the highest precision and stability.

Acknowledgements

The authors would like to acknowledge the support and continued encouragement of this project from Dr. Donald L. Hammond.

References

1. R. Holland, "Nonuniformly Heated Anisotropic Plates: I. Mechanical Distortion and Relaxation," *IEEE Trans. Sonics and Ultrasonics*, vol. SU-21, pp. 171-178, July 1974.
2. R. Holland, "Nonuniformly Heated Anisotropic Plates: II, Frequency Transients in AT and BT Quartz Plates," 1974 Ultrasonics Symposium Proceedings, pp. 593-598.
3. J. Kusters, "Transient Thermal Compensation for Quartz Resonators," *IEEE Trans. Sonics and Ultrasonics*, SU-23, pp. 273-276, July 1976.
4. E. EerNisse, "Quartz Resonator Frequency Shifts Arising from Electrode Stress," in *Proc. 29th Ann. Symp. on Frequency Control*, pp. 1-4, May 1975.
5. J. Kusters and J. Leach, "Further Experimental Data on Stress and Thermal Gradient Compensated Crystals," *Proc. IEEE*, vol. 65, pp. 282-284, Feb. 1977.

⁶ For a discussion of the measurement techniques, see Ref. 3 above, and C. Adams, J. Kusters and A. Benjaminson, "Measurement Techniques for Quartz Crystals, Proc. 22nd Ann. Symp. on Frequency Control, pp. 248-259, 1968.

⁷ W. Bond and J. Kusters, "Making Doubly Rotated Quartz Plates," presented at this symposium.



FIGURE 1 - Sealed and Unsealed 10 MHz TTC's.

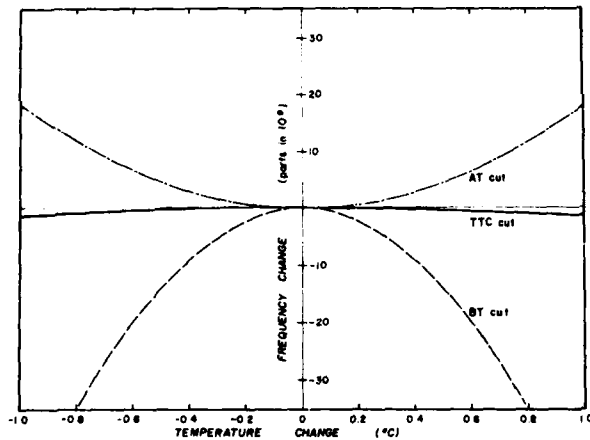


FIGURE 4 - Frequency-Temperature Characteristics @ 80°C.

CONTOUR	R_o	Q_o
1 DIOPTERS	55 Ω	1.25×10^6
2	70 Ω	1.45×10^6
3	97 Ω	1.22×10^6

TTC
10 MHz - 3rd OVERTONE

TABLE I. - Resistance and Q Data vs. Contour.

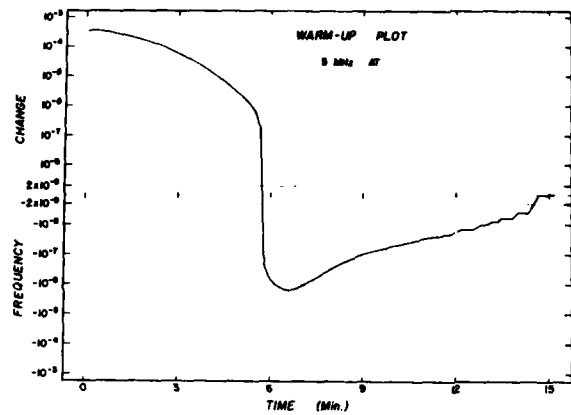


FIGURE 2 - Frequency Change - Fast Warm-up Oven - AT

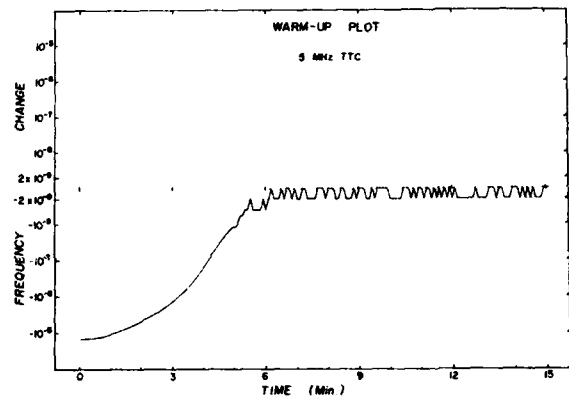


FIGURE 3 - Frequency Change - Fast Warm-up Oven - TTC

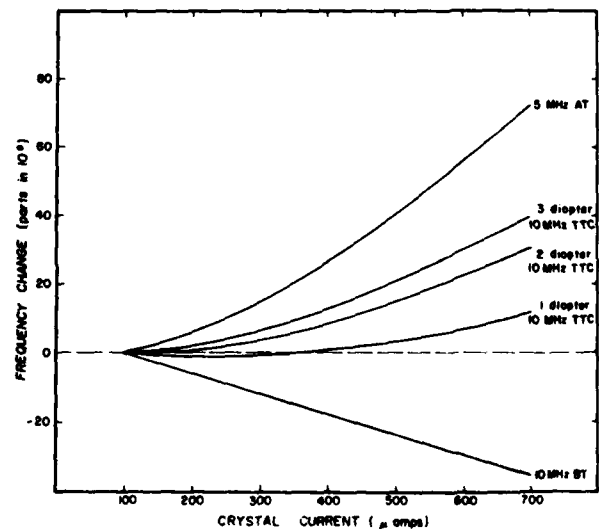
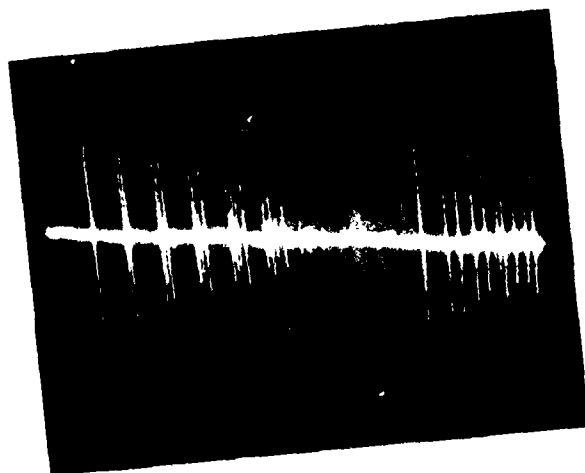
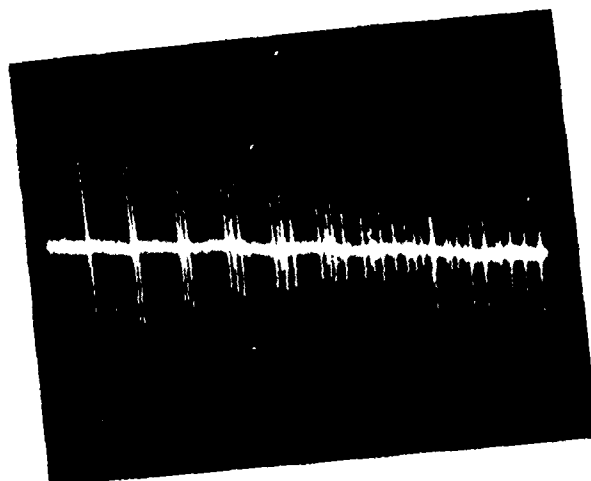


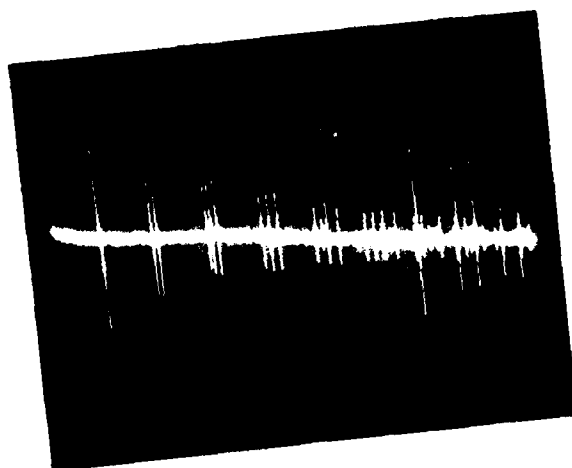
FIGURE 5 - Drive Level Dependence.



a) 1 DIOPTER



b) 2 DIOPTER



c) 3 DIOPTER

FIGURE 6 - Mode Spectra, 10 MHz, 3rd Overtone TTC Crystals. Horizontal Scale - 140 KHz/Div. Frequency Scan - 9.9 MHz to 11.3 MHz. Vertical Scale Uncalibrated - Proportional to Series Resistance.

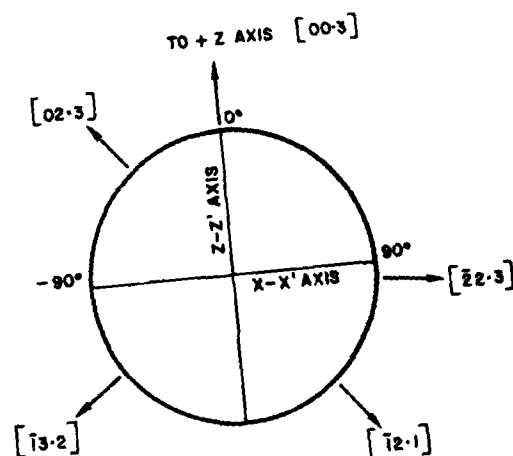


FIGURE 7 - Definition of Mounting Axes.

SUMMARY OF 2 DIOPTER TTC RESULTS

MOUNTING ANGLE	NO OF UNITS	R_o	Q_o	C_i	TURNOVER TEMPERATURE C°	AT TURNOVER	
						$T_f^{(2)} \times 10^9$	$T_f^{(3)} \times 10^{12}$
X-X'	7	67.3Ω	1.4×10^6	7.6 pf	71.3	-4.0	63
45°	11	69.1	1.6	7.7	72.9	-3.7	63
Z-Z'	15	73.3	1.3	7.7	73.5	-3.6	63

$\phi = 21.93^\circ$
 $\phi = 34.11^\circ$
 10MHz - 3/4 OVERTONE

TABLE II - Summary of 2 Diopter TTC Results.

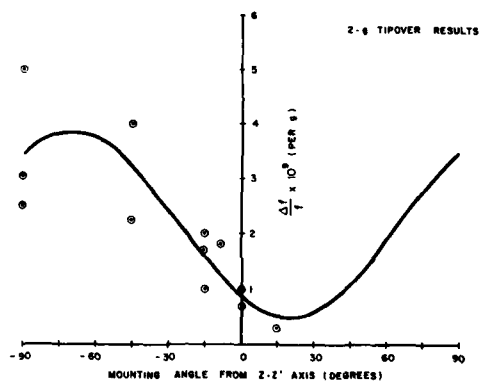


FIGURE 9 - '2-g' Tipover Results vs. Mounting Angle.

TTC vs CONVENTIONAL

BETTER

FAST WARMUP
 SECOND ORDER TEMPERATURE COEFFICIENT
 AMPLITUDE - FREQUENCY EFFECT
 ACTIVITY DIPS
 SHORT TERM STABILITY
 2-g TIPOVER

SAME

LONG TERM AGING
 SHUNT CAPACITANCE

WORSE

ANGULAR TOLERANCES
 Q_o
 R_o
 FREQUENCY SPECTRUM

FIGURE 12 - Comparison of TTC with Equivalent AT and BT Cuts.

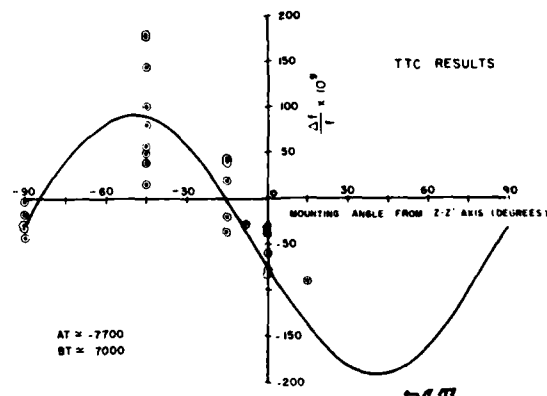


FIGURE 8 - Thermal Transient Compensation vs. Mounting Angle

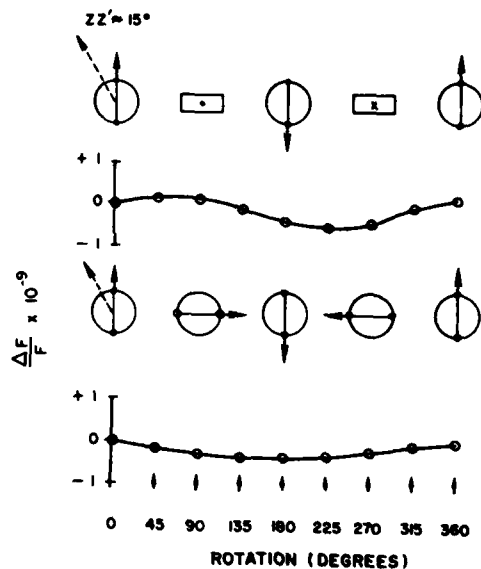


FIGURE 10 - '2-g' Tipover Results @ 15° From Z-Z' Axis.

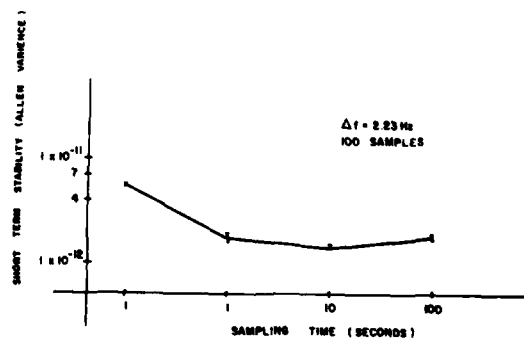


FIGURE 11 - Short Term Stability - 2 Diopter TTC.

THE FORCE-FREQUENCY EFFECT IN DOUBLY ROTATED QUARTZ RESONATORS

A. Ballato,[†] E. P. EerNisse* and T. Lukaszek[†]

[†]US Army Electronics Technology & Devices Laboratory (ECOM)
Fort Monmouth, New Jersey 07703

*Sandia Laboratories
Albuquerque, New Mexico 87115

Summary

Precision frequency control requirements for future digital communication systems require improved crystal resonator performance in a number of aspects. Accordingly, the potential of doubly rotated quartz cuts has begun to be explored. In the neighborhood of the SC-cut ($YX\omega l$) ϕ/θ , $\phi = 21.9^\circ$, $\theta = +33.9^\circ$, a variety of effects having their bases in nonlinear elasticity have been shown, or are predicted, to be improved with respect to corresponding AT-cut values. The static frequency-temperature behavior also shows improvement.

This paper concerns the force-frequency effect, thus far not investigated in any detail for doubly rotated quartz plate vibrators. It relates initial stress produced by mounting supports to resonance frequency changes; it contributes to long-term aging, and is related to the frequency excursions produced in shock and vibration environments.

In-plane diametric forces applied to the periphery of vibrating plates produce frequency changes that depend upon the azimuth angle ψ in the plane of the plate. For the IT-cut at $\phi = 19.1^\circ$, the maximum value was found previously to be only about one-third that of the AT-cut. This points to a reduced coefficient at the SC-cut as well. Measurements of the force frequency effect coefficients have now been extended to doubly rotated quartz plates. Also given are charts of the mode spectra in the region of the thickness modes, and the modal temperature coefficients.

The force coefficient data are compared with theoretically predicted values obtained from a variational principle applied to an anisotropic disc supported at two diametric points. This analysis departs from previous treatments in two major respects: 1) the isotropic stress pattern is replaced by anisotropic stress; 2) the elastic problem is treated for triclinic symmetry, rather than monoclinic symmetry. These investigations verify the superiority of doubly rotated plates with respect to the force-frequency effect and provide further motivation for their continued development and utilization.

Introduction

Precision frequency control requirements for digital communication and position location systems currently undergoing development make it imperative that crystal resonator performance be improved in a number of aspects. Accordingly, the potential of doubly rotated quartz cuts has begun to be explored.¹⁻¹⁰ For cuts on the upper zero temperature coefficient locus in general ($\theta = +34^\circ$), and in the neighborhood of the SC-cut in particular ($\phi = 21.9^\circ$, $\theta = +33.9^\circ$), a variety of effects having their bases in nonlinear elasticity have been shown, or are predicted, to be reduced below the corresponding AT-cut values. In addition, the static frequency-temperature behavior shows some improvement.

Among the nonlinear effects of interest are:

- Force-frequency
- Acceleration - frequency
- Resonance amplitude - frequency
- Intermodulation
- Mode coupling - activity dips
- Dynamic thermal - frequency
- Film stress - frequency

Some of these have received, or are receiving, theoretical and/or experimental treatments; this paper is principally concerned with the force-frequency effect which has thus far not been investigated in any detail for doubly rotated quartz plate vibrators. This effect relates the initial stress produced by the mounting supports to resonance frequency changes; it contributes to long-term aging and is also related to the frequency excursions produced in shock and vibration environments.

In-plane diametric forces applied to the periphery of vibrating plates produce frequency changes (order 10^{-7} per gram) that depend upon the azimuth angle ψ in the plane of the plate. If ψ is measured from the X_1'' axis, then it is found experimentally¹¹ that for the AT-cut the effect is zero at ψ values of 60° and 120° . For the IT-cut at $\phi = 19.1^\circ$, Ballato¹ found the zeros to occur at $\psi = 85^\circ$ and 163° with a maximum value about one-third that of the AT-cut. This points to a reduced coefficient in the SC-cut as well.

In this paper we extend the force-frequency effect measurements to doubly rotated quartz plates on the upper zero temperature coefficient locus, concentrating on the SC- and FC-cuts because of their technological significance. Also given are charts of the mode spectra in the region of the thickness modes and the modal temperature coefficients.

The force coefficient data are compared with theoretically predicted values obtained from a variational principle applied to an anisotropic disc supported at two diametric points. This analysis departs from previous treatments^{2,5,12-15} in two major respects: 1) the isotropic stress pattern is replaced by the more accurate anisotropic stress; 2) the elastic problem is treated for the general triclinic symmetry, rather than the monoclinic symmetry appropriate to rotated-Y-cuts.¹⁶

These investigations verify the predicted superiority of doubly rotated quartz plates over the conventional AT-cut, with respect to the force-frequency effect, and provide further motivation for their continued development and utilization.

Doubly Rotated Crystal Plates

Doubly rotated crystal plates are the most general kind of one-dimensional thickness-mode vibrator. The orientation is uniquely specified by two angles-- ϕ and θ . Following the usual convention,¹⁷ the orientation with respect to the crystallographic axes is described as $(YX\omega l)\phi/\theta$. Examples of singly and doubly rotated cuts are shown in Figure 1, along with the angles.³ Also shown in the figure is the locus of zero first

order temperature coefficient (ZTC) for quartz resonator plates. In quartz, the first rotation lowers the apparent symmetry from trigonal to monoclinic; the second rotation further lowers it to triclinic.

Static Frequency-Temperature Behavior

For quartz plates on the upper ($\theta > 0$) ZTC locus, the static frequency-temperature (f-T) curve exhibits a cubic behavior. The AT-cut is the classical example. Whereas the inflection temperature (the temperature half-way between turnover points) occurs at room temperature for AT-cuts, this temperature increases steadily with increasing angle ϕ , becoming 48°C at the FC-cut, 74°C at the IT-cut, 95°C at the SC-cut, and 157°C for the rotated-X-cut. Typical f-T curves for the SC-cut are shown in Figure 2 for the c- and b-modes. Dots represent measured points. For the b-mode (the faster, quasi-shear mode, classed as undesired), the first order temperature coefficient is $-25.1 \times 10^{-6}/K$. The c-mode (slower, quasi-shear, desired mode) curve was fit by quintic least-squares to yield the following coefficients at 25°C:

$$\begin{aligned} a &= 0.38 \times 10^{-6}/K \\ b &= -11.4 \times 10^{-9}/K^2 \\ c &= 26.8 \times 10^{-12}/K^3 \\ d &= 136. \times 10^{-15}/K^4 \\ e &= 175. \times 10^{-18}/K^5 \end{aligned}$$

The SC-cut is seen to have its inflection temperature around 100°C, so that it would normally be operated around its lower turnover temperature, where the upper turnover is used for the AT-cut. Compared with the AT-cut, the SC-cut is also flatter, so a given temperature control will correspond to a smaller frequency deviation.

Mode Spectrographs³

A wide band mode spectrograph is shown in Figure 3 for an SC-cut crystal. The modes, in order of frequency, are denoted as $m(M)$, where m is the mode type (a, b, or c) and M is the order of the harmonic. The sequence shown is $c(1)$ $b(1)$ $a(1)$ $c(3)$ $b(3)$ $c(5)$ $b(5)$ $a(3)$ and $c(7)$.

The thickness-shear mode TS_1 at cutoff corresponds to the c-mode; the thickness-twist TT_2 at cutoff corresponds to mode b, and mode a is the thickness-stretch mode, TE. The spacings and amplitudes measured agree closely with those calculated last year.³

In Figure 4 the spectrum in the vicinity of the c-mode is shown; a narrowband plot about the b-mode resonance is given in Figure 5. From Figures 3, 4, and 5 one sees how very clean the spectrum is, even for the harmonic modes. The flat SC-plate used for this experiment had the following measurements:

plate diameter $\phi_a = 14.18$ mm;
electrode diameter $\phi_e = 5.0$ mm (keyhole pattern);
mass loading (plateback) $\mu = 1.8\%$;
c-mode fundamental frequency = 5.937 MHz.

Although one would suspect that the SC orientation, because of its lower symmetry, would have a more complicated unwanted mode spectrum, it appears that energy trapping^{18,19} can be applied readily to these plates, although the optimum electrode shape²⁰ and plateback relations are not available at the present time.

Mathematical Modeling

An overview of the present theoretical work, compared to past treatments, is shown in Figure 5.

Past theoretical analyses^{2,5,12,13} of static mechanical stress bias effects, in general, and of force sensitivities, in particular, have been two-step calculations. First, linear elastic solutions for the distribution of static mechanical stress bias in the resonator blank were obtained, assuming that quartz is isotropic. These isotropic static solutions for the stress at the blank center were then used to calculate resonant frequency shifts in non-linear wave propagation calculations which included the correct anisotropy of quartz and third-order elastic constant effects.

The present work is a two-step calculation where an attempt is made for a better solution to the static problem. The calculus of variations (essentially the Rayleigh-Ritz method) is used to find approximate solutions to the anisotropic static stress problem. The approximate static solution for the stress at the blank center is used in a non-linear elastic wave propagation code to calculate resonant frequency shifts.

The present theoretical results for the singly-rotated AT- and BT-cuts provide a much improved comparison with published experimental results than the earlier theoretical results. The theory is used to calculate force sensitivity coefficients for the doubly-rotated cut family containing the important AT-, IT-, FC-, and SC-cuts. The results provide the crystal designer with the appropriate azimuthal angle to mount the resonator on a two-point mount for minimum force sensitivity.

A. Theory

One calculational approach that has been useful on numerous occasions for calculating stress patterns in static and vibrating elastic material bodies is the calculus of variations.^{21,22} The method is approximate although the closeness to which the approximate solution can be brought to the actual solution is a matter of degree depending on the choice of trial (basis) functions, available computer size, and patience. The method for static problems amounts to formulating the total elastic stress-strain energy stored in a given body for the given boundary conditions and trial functions and then adjusting the trial functions to minimize the stored elastic energy. The approximations obtained with the method are somewhat better for the elastic energy values than the stress distributions, but sufficient accuracy of the stress distributions can be obtained for practical considerations.

The total stored elastic energy L is given by

$$L = \frac{1}{2} \iiint_V dV S_{\lambda\mu} C_{\lambda\mu} - \iint_S dS F_i U_i \quad (1)$$

Here V and S are the resonator blank volume and surface, $C_{\lambda\mu}$ is the elastic stiffness tensor in engineering notation, F_i is the distribution of force per unit area acting on the surface, U_i is the elastic displacement vector. We use a cartesian coordinate system x_i for the plate, and

$$S_{ij} = (U_{i,j} + U_{j,i})/2 \quad (2)$$

Here λ and μ run 1-6, and i and j run 1-3. S_{ij} is

related to S_{ij} by the conventional²¹ relations between engineering and tensor notations.

The present calculations treat a circular resonator blank of diameter d and thickness τ described by the IRE standard¹⁷ notation $(YXw\ell)\phi/\theta$. We consider only contoured or energy trapped resonator designs where the vibrational acoustic energy is restricted to the vicinity of the blank center. Contouring effects are ignored for the static stress distribution calculation, however, so that the much simpler problem of a flat circular plate can be solved for the static stress distribution. This simplification is warranted because the thin resonator blanks used for thickness shear resonators allow the assumption that the thickness x_2 (or y) dimension is small enough for a plane stress problem.

In the case of plane stress, λ and μ run 1, 3, 5 and 1 and j run 1, 3, in Equations 1 and 2 (x_2 is blank thickness direction, x_1 is ℓ , x_3 is w in the standard notation). Also, $C_{\lambda\mu}$ is replaced by $\gamma_{\lambda\mu}$, the planar elastic stiffness coefficients expressed in the plate coordinate system. Hence:

$$L = \frac{\tau}{2} \iint dx_1 dx_3 dV S_{\lambda\mu} S_{\lambda\mu}^v - \iint dS F_1 U_1 \quad (3)$$

The variational method involves substituting a linear superposition of trial functions for U_1 in Equations 2 and 3, carrying out the integrals in Equation 3, and minimizing the resulting expression (differentiating the expression with respect to a given coefficient and setting that equal to zero) with respect to the coefficients of the trial functions. The problem then becomes a linear algebra problem in the coefficients. The choice of trial functions must be such that they represent a pointwise complete set over V and S . If the trial functions already satisfy some aspect of the problem such as the differential equation or boundary conditions, the number of trial functions needed for adequate convergence is small. For the present case, we take the easy way out and choose a simple power series expansion of U_1 and rely on the computer to handle large numbers of trial functions. The trial functions chosen for the two-point problem are

$$U_1 = \sum_{m,n,p,q=0}^{MNPQ} \left\{ A_{mn} x_1^{2m+1} x_3^{2n} + B_{pq} x_1^{2p} x_3^{2q+1} \right\} \quad (I)$$

and

$$U_3 = \sum_{pqtu}^{PQTU} \left\{ B_{pq} x_1^{2p+1} x_3^{2q} + D_{tu} x_1^{2t} x_3^{2u+1} \right\} \quad (II)$$

The value for F_1 is set to equal

$$F_1 = F/(\delta\tau), \quad (4)$$

where F is the inwardly acting force applied to the opposite ends of a blank diameter, τ is the plate thickness, and δ is some length dimension which is small relative to the blank circumference (point force). All the integrals can be carried out in Equation 3, the surface integral being treated in the limit of a point force ($\delta \rightarrow 0$). After differentiation with respect to A_{mn} , B_{pq} , and D_{tu} , the resulting linear algebra problem²² has a solution for A_{mn} , B_{pq} , and D_{tu} which scales with F . Thus the static stress distribution can be solved for any known $\gamma_{\lambda\mu}$. A computer code for the linear algebra problem was written

which included rotation of the quartz elastic tensor to obtain $\gamma_{\lambda\mu}$ and arbitrary selection of M , N , P , Q , T , and U . From symmetry arguments, one is led to use groups of the trial functions, adding new groups until satisfactory convergence is obtained. The groups are defined by $(0,R)$, where the group includes all pairs of (m,n) , (p,q) , and (t,u) with the first member increasing from zero to R in steps of one while the second member decreases from R towards zero in steps of one, e.g., $(0,3)$, $(1,2)$, $(2,1)$, $(3,0)$ make up the family $(0,3)$.

The resulting solution for the static stress at the plate center using the published c_{ijk}^P constants for quartz²³ is incorporated into a previously described computer code^{2,5} which calculates the resonant frequency shift caused by elastic nonlinearities (third-order elastic constant effects). We use here the definition for the force sensitivity coefficient K_f defined as¹⁶

$$\frac{\Delta f}{f} = K_f \frac{FNo}{d\tau} \quad (5)$$

where No is the frequency constant. The plate diameter is $d = \phi_a$. Units are K_f in $m^2 \text{sec}^2/N$, No in $m^2 \text{sec}$, d and τ in m , and f in sec^{-1} or Hz . K_f is positive if frequency increases upon application of a compressive force.

The direction of the applied forces F is important because of the anisotropy of the quartz nonlinear elastic problem and because of the quartz anisotropy in the static stress problem (the latter is ignored in earlier isotropic static solutions). We choose to follow the earlier experimental work by presenting results for the K_f of a given $(YXw\ell)\phi/\theta$ cut as a function of azimuthal angle ψ . The azimuthal angle is measured in right-hand convention about x_2 from x_1 (or ℓ) in the plane of the plate (positive angle going from x_1 to $-x_3$; our x_1 , the plate coordinate system, sometimes is referred to as x_1^1).

B. Results

The program was first tested for isotropic blanks. The solution at the center of the plate converges rapidly (four significant figures) to that found analytically,²⁴ i.e., $6F/\pi d\tau$ compression along the diameter aligned with (inwardly directed) F and $2F/\pi d\tau$ tension along the diameter perpendicular to F .

Calculations were carried out for the AT- and BT-cuts since there exist large amounts of force sensitivity data and considerable discrepancies between the data and earlier theoretical results for these cuts. The results for the zero temperature coefficient thickness-shear c-mode are shown in Figure 7 as K_f vs ψ . The "isotropic assumption" in Figure 7 is the result obtained in the present computer codes when the isotropic solution ($-6F/\pi d\tau$, $2F/\pi d\tau$) is used for the static stress pattern at the plate center. This result, is numerically equivalent to the results of Lee, et al.¹³ Note the discrepancies between the isotropic assumption results and the experimental summary provided by Ratajski;¹⁶ namely, the difference in value for K_f at $\psi = 0$ for the AT-cut, and the complete failure to predict the dip in the BT-cut results. As seen in Figure 7, the present calculations using the variational treatment are quantitatively accurate at $\psi = 0$ for the AT-cut and predict a dip for the BT-cut. The calculations leading to Figure 7 involved using all the families up to and including $(0,5)$, making 21 values each for A_{mn} , B_{pq} , and D_{tu} for a 63×63 linear algebra problem. The addition of the $(0,5)$ family only changed the numerical answer for the stress at the plate center in the

third significant figure, so some idea of convergence is provided.

Figure 8 contains present results for the thickness shear (mode c) IT-cut and previously published experimental results.¹ The qualitative features of the experimental data are reproduced by both the isotropic assumption results and the full variational result. In view of the results in Figure 7, where more experimental data exist, we consider the variational result to be the more accurate. The quantitative discrepancy between experiment and theory is not as serious as might appear in Figure 8 because there is a factor-of-two scale expansion between Figures 8 and 7 to account for the generally smaller K_f values of doubly rotated cuts.

A series of calculations has been carried out for the zero temperature coefficient branch (mode c) subset within $(YX\omega_2)\phi/\theta$: $\phi = 0^\circ$ (AT-cut); $\phi = 10^\circ$; $\phi = 15^\circ$ (FC-cut); $\phi = 19^\circ$ (IT-cut); $\phi = 22^\circ$ (SC-cut); $\phi = 30^\circ$ (rotated-X-cut). The results for K_f vs ψ are shown in Figure 9.

The family of curves for K_f vs ψ in Figure 9 are the general theoretical results for this paper. The curves should be useful in choosing the optimum location of bonding pads for two-point mounts to minimize the effects of mounting stresses and some effects of acceleration.

In Table 1 are given the computed angles ψ at which K_f is zero, as function of ϕ . Table 2 lists the computed locations and values of the extrema of K_f as function of ϕ .

TABLE 1. ZEROS OF K_f

ϕ°	ψ°
0	64.7, 115.3
10	68.5, 125.2
15	74.8, 148.8
19.1	79.3, 163.1
21.9	81.6, 171.9
30	79.3, 184.3

TABLE 2. LOCATION AND VALUES OF K_f EXTREMA

ϕ°	ψ°	K_f (max)	ψ°	K_f (min)
0	0	24.5	90	-11.5
10	17.6	21.4	93.2	-10.8
15	26.0	18.2	98.8	-8.9
19.1	36.6	15.3	118.1	-10.1
21.9	44.3	14.7	131.6	-14.3
30	55.6	14.7	130.0	-23.9

(K_f in 10^{-15} m·s/N)

Experimental Confirmation

Apart from the IT-cut results obtained in 1960,¹ no experimental results were available for comparison with the theoretically predicted force-frequency curves of doubly rotated cuts of quartz, shown in Figure 9. In order to obtain these results, the experimental apparatus of Figure 10 was constructed. It contains provisions for the precise rotation of the angle ψ by means of a vacuum chuck for holding the crystal (and maintaining strict crystal vertical alignment), and high-ratio reduction gear. Five mil wires were bonded to the crystal and served for the electrical connections. To the left in the figure, the connections are brought out to a Crystal Impedance Meter (RFL model 459 = TS-330 with low drive

modification). The frequency is recorded from a counter.

Micrometer adjustments are available for assuring accurate alignment of the various portions of the jig. Force application was made by a movable rod on which calibrated masses were applied; the mass of the rod was taken into account.

The crystals considered here had orientations $(YX\omega_2)\phi/\theta$, with $\phi = 10^\circ, 15^\circ$, and 21.9° , and θ such that the units had zero temperature coefficients ($\theta = +34^\circ$). All crystal units were provisionally scored near the X_3'' axis (projection on the Z, or optic axis). The true location of X_3'' with respect to the score mark was later determined for each crystal to within 1° by a conoscope, and enabled ψ to be accurately known. (This procedure was also applied to the still-extant IT-cut crystal that was described in Reference 1. The new measurement disclosed that the true X_3'' axis as seen in the conoscope was a full 19° in error with respect to the score mark on the crystal. Thus, the curve in Figure 18 of Reference 1 ought to be translated to the right with respect to the graph axes, so that the zeros occur at 85° and 163° . This finding was a welcome resolution of a disturbing discrepancy between theory and experiment!)

Measurements were made in ψ intervals of 10° . Three readings were taken at each ψ . First, the frequency of the unloaded vibrator was measured. Then the weight was lowered gently, and the loaded frequency was recorded. Finally, the frequency with the weight removed was measured. The first and third frequencies were, in all cases, at most one or two Hz apart. K_f was then calculated from equation (5) using the measured values of frequency change Δf with applied weight, plate diameter and thickness, force applied, and crystal frequency constant No. For the three cuts considered, the frequency constants are¹⁰

$\phi = 10^\circ$,	No = 1690 m/s
15°,	1726
21.9°,	1797.

Figures 11, 12, and 13 show the experimental curves for the c-mode cuts at $\phi = 10^\circ, 15^\circ$, and 21.9° , respectively. Included in each figure is a curve, shown dashed, of the theoretical result obtained from the variational procedure outlined earlier. The solid curves represent experimental results averaged over a group of units, and also over a number of runs on each unit by three experimenters. The error bars represent data extremes encountered.

Data scatter is worst for the $\phi = 10^\circ$ units, and a fully satisfactory explanation for this cannot at present be given. In all cases it was observed that the unloaded frequencies appeared to vary with ψ , indicating, e.g., a stray capacitance effect due to the fixture. A modification of the mounting arrangement to reduce this effect produced the solid curve shown in Figure 13 without the error bars. As may be seen in the figure, the modification improves the agreement of the experimental curve with the theoretical. In Figures 11, 12, and 13, the agreement between experiment and theory is generally quite good, especially when the smallness of the effect is borne in mind. The overall magnitudes predicted and observed agree well, as do the general features in each case. The symmetry about $\psi = 0^\circ$ and 90° observed in the AT- and BT-cut curves (cf. Figure 7) is seen to be absent in the doubly rotated cut curves of Figures 8, 11, 12, and 13; this is because the diagonal axis of symmetry, which lies in the plane of rotated-Y-cut plates, is out of the plane for cuts having $\phi \neq 0^\circ$.

A comparison of Figures 12 and 13 with Figure 7 reveals that the peak-to-peak excursions of K_f with ψ for the FC- and SC-cuts is about one half that of the AT-cut. This indicates a reduced sensitivity of these cuts to mechanical shock. Further comparison between Figures 12 and 13 discloses that the average force coefficient

$$\langle K_f \rangle = \frac{1}{\pi} \int_0^{\pi} K_f(\psi) d\psi \quad (6)$$

for the FC-cut is positive and relatively large (albeit lower than for the AT-cut), whereas for the SC-cut it is very small (ideally zero). This feature stems from the SC-cut definition as the ZCT orientation for which planar stress produces no frequency change.

Figure 14 gives the results of a preliminary set of measurements of the force sensitivity for the b-mode of the SC-cut. This mode has frequency constant¹⁰

$$N_0 = 1977 \text{ m/s},$$

and a temperature coefficient of approximately -25 or -26 ppm/K. Because of the large temperature coefficient, the force coefficient measurements are not as easy to make as for the c-mode. The results reported in Figure 14 should be regarded as tentative; they disclose, however, that K_f exhibits no zeros along the ψ axis.

Conclusions

Doubly rotated quartz resonators have been considered as to frequency-temperature behavior, mode spectrum content, and primarily with respect to the force-frequency effect. It was found that the f - T curve shifts upward in temperature with increasing ϕ angle, and becomes flatter. The mode spectrum is complicated by the presence of all three thickness modes, but their anharmonic overtones may be very adequately suppressed by means of energy trapping.

In the force-frequency portion, the calculus of variations has been used to obtain solutions for the static stress distributions caused in circular quartz resonator blanks by forces acting along a diameter. The resulting static stress distributions have been used to predict the force sensitivity of the resonant frequency of thickness shear resonators. The present results are shown to agree more favorably with experiment than results obtained by previous theories where the quartz was assumed to be isotropic for the static stress solution. General results are given for the doubly-rotated family (YXw ϕ) ϕ /6 along the technologically important zero temperature branch which contains the AT-, FC-, and SC-cuts. The results will be useful for choosing the location of bonding pads for two-point mounts so that mounting stress effects can be minimized.

Acknowledgment

The work of E. P. EerNisse was supported by the United States Energy Research and Development Administration (ERDA) under Contract No. AT(29-1)789.

We take this opportunity to thank Mr. Richard S. Tilton of ECOM for his enthusiastic assistance; Prof. Raymond Besson, Université de Besançon for generously providing the SC-cuts used in these experiments; and Mrs. Anita Earle for her expert typing.

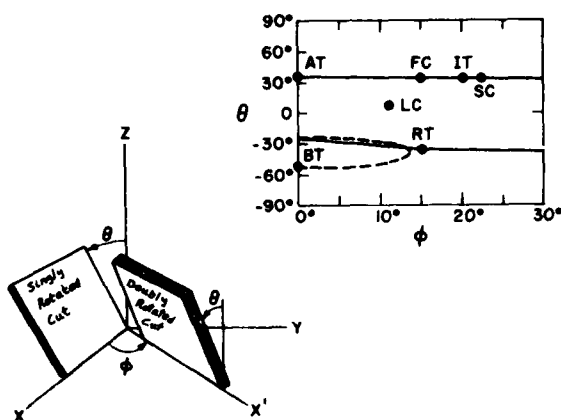
References*

1. A. D. Ballato; "Effects of Initial Stress on Quartz Plates Vibrating in Thickness Modes," Proc. 14th AFCS, May-June 1960, pp. 89-114.
2. E. P. EerNisse; "Quartz Resonator Frequency Shifts Arising from Electrode Stress," Proc. 29th AFCS, May 1975, pp. 1-4.
3. A. Ballato and G. J. Iafrate; "The Angular Dependence of Piezoelectric Plate Frequencies and Their Temperature Coefficients," Proc. 30th AFCS, June 1976, pp. 141-156.
4. J. Kusters; "Transient Thermal Compensation for Quartz Resonators," IEEE Trans. Sonics Ultrason., Vol. SU-23, No. 4, July 1976, pp. 273-276.
5. E. P. EerNisse; "Calculations on the Stress Compensated (SC-Cut) Quartz Resonator," Proc. 30th AFCS, June 1976, pp. 8-11.
6. R. L. Filler and J. R. Vig; "The Effect of Bonding on the Frequency vs. Temperature Characteristics of AT-Cut Resonators," Proc. 30th AFCS, June 1976, pp. 264-268.
7. J. A. Kusters and J. G. Leach; "Further Experimental Data on Stress and Thermal Gradient Compensated Crystals," Proc. IEEE, Vol. 65, No. 2, Feb. 1977, pp. 282-284.
8. J. -J. Gagnepain, J. C. Ponçot, and C. Peugeot; "Amplitude-Frequency Behavior of Doubly Rotated Quartz Resonators," these Proceedings.
9. J. A. Kusters, C. A. Adams, and H. Yoshida; "Thermal Transient Compensated Crystal Cuts--TTC's; Further Developmental Results," these Proc.
10. A. Ballato; "Doubly Rotated Thickness Mode Plate Vibrators," in *Physical Acoustics: Principles and Methods*, (W. P. Mason and R. N. Thurston, eds.), Vol. 13, Chap. 5. Academic Press, New York, 1977, pp. 115-181.
11. E. A. Gerber and M. H. Miles; "Reduction of the Frequency-Temperature Shift of Piezoelectric Resonators by Mechanical Stress," Proc. IRE, Vol. 49, No. 11, November 1961, pp. 1650-1654.
12. R. W. Keyes and F. W. Blair; "Stress Dependence of the Frequency of Quartz Plates," Proc. IEEE, Vol. 55, No. 4, April 1967, pp. 565-566.
13. P. C. Y. Lee, Y. S. Wang, and X. Markenscoff; "Elastic Waves and Vibrations in Deformed Crystal Plates," Proc. 27th AFCS, June 1973, pp. 1-6.
14. P. C. Y. Lee and D. W. Haines; "Piezoelectric Crystals and Electro-elasticity," in *R. D. Mindlin and Applied Mechanics*, (G. Herrmann, ed.), Pergamon Press, New York, 1974, pp. 227-253.
15. P. C. Y. Lee, Y. S. Wang, and X. Markenscoff; "High Frequency Vibrations of Crystal Plates Under Initial Stresses," J. Acoust. Soc. Amer., Vol. 57, No. 1, Jan. 1975, pp. 95-105.

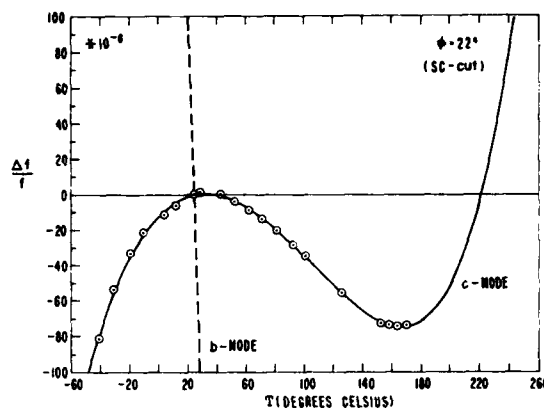
*AFCS: Annual Frequency Control Symposium, US Army Electronics Command, Fort Monmouth, NJ 07703.

16. J. M. Ratajski; "Force-Frequency Coefficient of Singly-Rotated Vibrating Quartz Crystals," IBM J. Res. Dev., Vol. 12, No. 1, January 1968, pp. 92-99.
17. "Standards on Piezoelectric Crystals, 1949," Proc. IRE, Vol. 37, No. 12, December 1949, pp. 1378-1395. (IEEE Standard No. 176).
18. W. Shockley, D. R. Curran, and D. J. Koneval; "Trapped-Energy Modes in Quartz Filter Crystals," J. Acoust. Soc. Amer., Vol. 41, No. 4 (Part 2), April 1976, pp. 981-993.
19. T. J. Lukaszek and H. P. Wasshausen; "New Design Information for Filter Crystals in the 3.9-10 Mc/s Frequency Range," Proc. IEEE, Vol. 53, No. 11, November 1965, p. 1772.
20. R. D. Mindlin; "Optimum Sizes and Shapes of Electrodes for Quartz Resonators," J. Acoust. Soc. Amer., Vol. 43, No. 6, June 1968, pp. 1329-1331.
21. R. Holland and E. P. EerNisse; Design of Resonant Piezoelectric Devices, Research Monograph No. 56, M.I.T. Press, Cambridge, MA, 1969.
22. S. H. Gould; Variational Methods for Eigenvalue Problems, Univ. Toronto Press, Toronto, and Oxford Univ. Press, London, 1966.
23. H. J. McSkimin, P. Andreatch, Jr., and R. N. Thurston; "Elastic Moduli of Quartz versus Hydrostatic Pressure at 25° and -195.8°C," J. Appl. Phys., Vol. 36, No. 5, May 1965, pp. 1624-1632.
24. S. Timoshenko and J. N. Goodier; Theory of Elasticity, McGraw-Hill, New York, 1951.

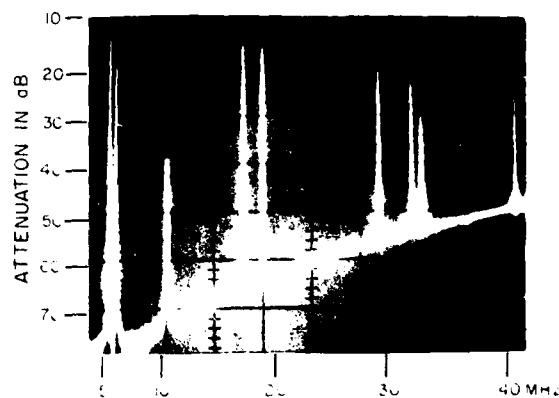
DOUBLY ROTATED QUARTZ CUTS



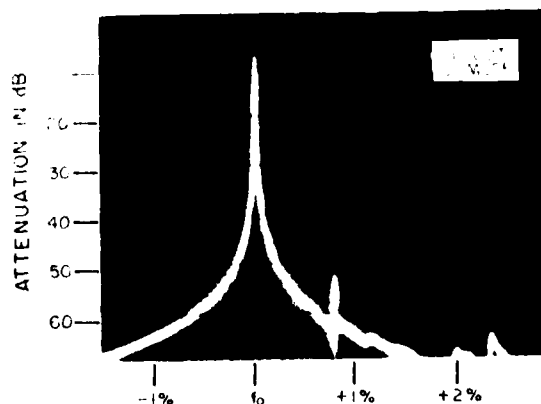
1. Conventions for specifying plate orientations; locus of zero temperature coefficient in quartz.



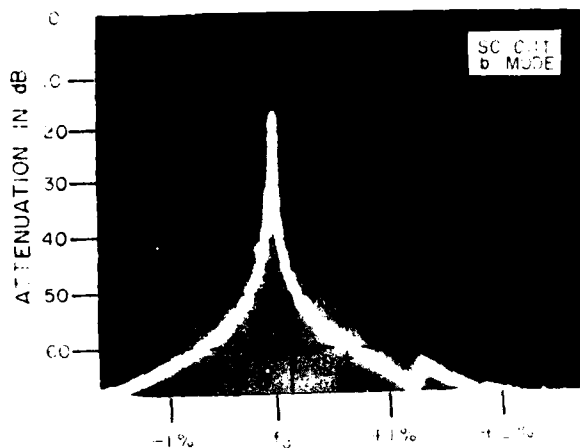
2. Frequency-temperature behavior of an SC-cut.



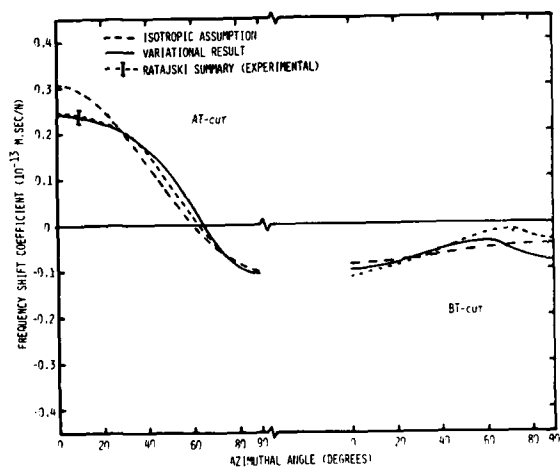
3. Wideband mode spectrograph of an SC-cut.



4. Narrowband spectrograph of the SC-cut c mode.



5. Narrowband spectrograph of the SC-cut b mode.

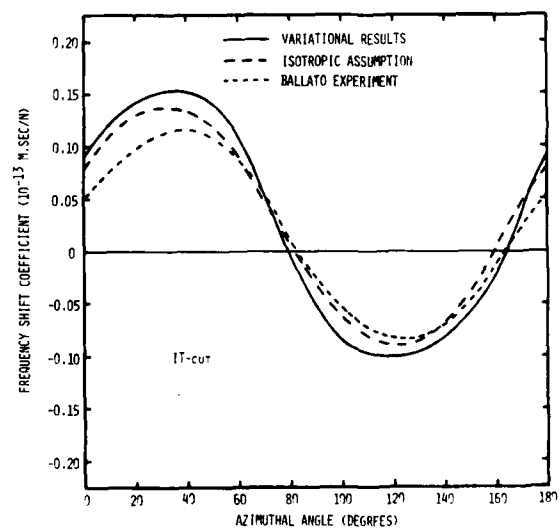


7. Force-frequency plots for AT- and BT-cut quartz.

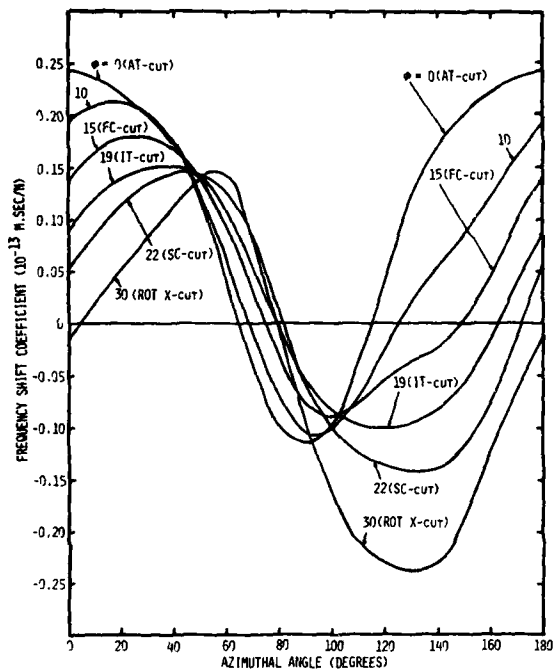
MATHEMATICAL MODELING OF FORCE-FREQUENCY EFFECT

- PRIOR
- ISOTROPIC INITIAL STRESS FIELD ASSUMED.
 - ANALYSIS LIMITED TO ROTATED-Y-CUTS $(YX\theta)\theta$.
- PRESENT
- ANISOTROPIC INITIAL STRESS FIELD OBTAINED BY CALCULUS OF VARIATIONS.
 - SOLUTION APPLIED TO DOUBLY ROTATED CUTS $(YX\omega)\theta/\theta$.

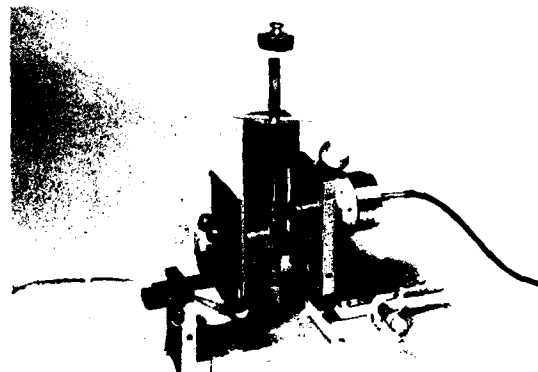
6. Outline of mathematical analyses.



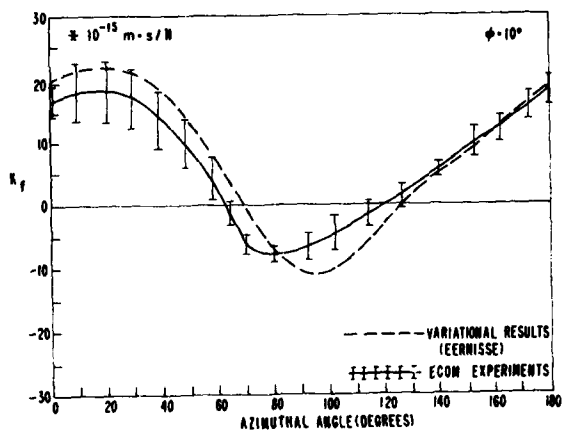
8. Force-frequency plots for IT-cut quartz.



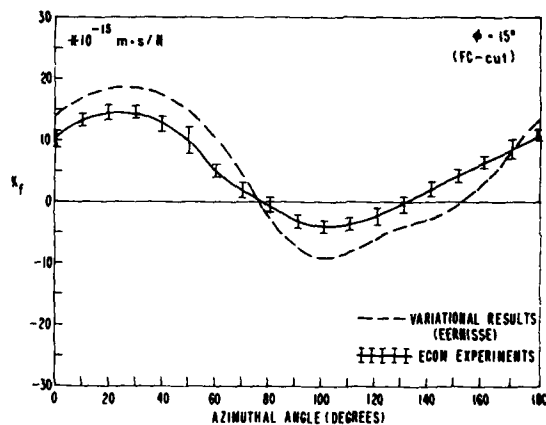
9. Composite of force-frequency plots for quartz cuts.



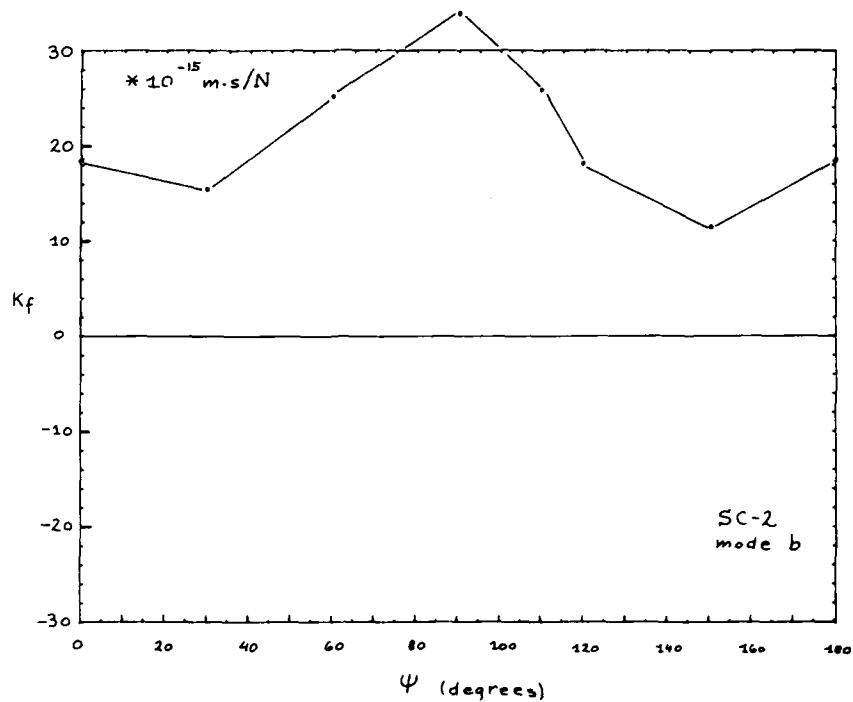
10. Photograph of force-frequency apparatus.



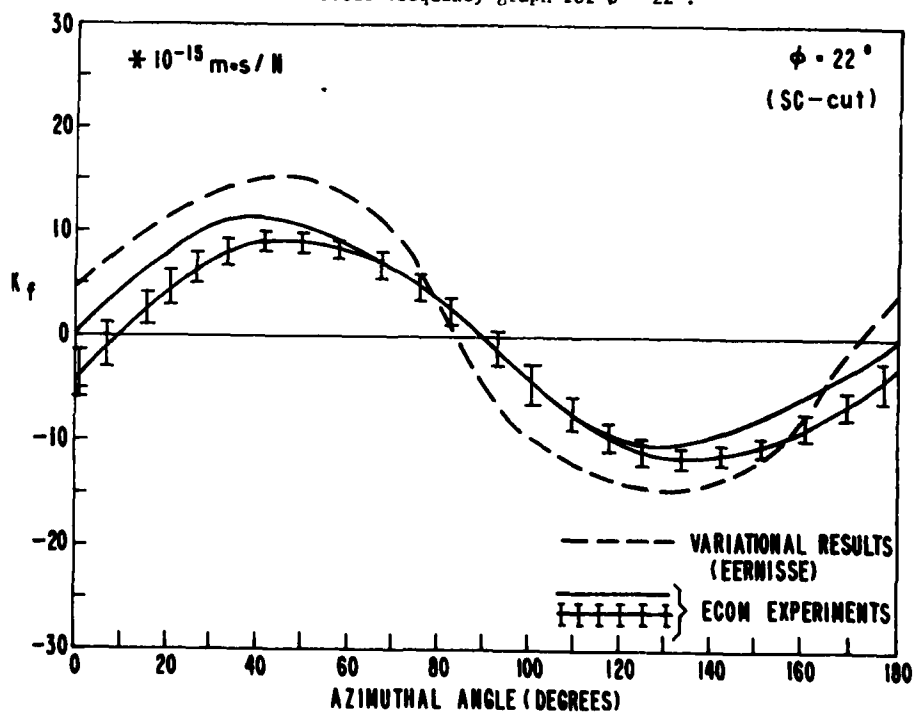
11. Force-frequency graph for $\phi = 10^\circ$.



12. Force-frequency graph for $\phi = 15^\circ$.



13. Force-frequency graph for $\phi = 22^\circ$.



14. Force-frequency graph for the SC-cut b mode.

AMPLITUDE - FREQUENCY BEHAVIOR OF
DOUBLY ROTATED QUARTZ RESONATORS

J.-J. Gagnepain
Laboratoire de Physique et Métrologie des Oscillateurs du C.N.R.S.
32 avenue de l'Observatoire 25000 Besançon France

J.-C. Ponçot
Ecole Nationale Supérieure de Chronométrie et de Micromécanique
route de Gray 25000 Besançon France
and
C. Pegeot
Quartz et Electronique
1 rue d'Anjou 92502 Asnières France

Summary

The amplitude-frequency effects of quartz crystal resonators are due to different physical mechanisms. The main phenomenon is related with the crystal non linear elastic properties : elastic coefficients of the 3rd and 4th order, non linear propagation equation. But thermal effects can induce also a frequency change under a variation of the drive level. Two kinds of thermal behavior are to be distinguished : the quasi-static thermal effects related with the temperature coefficients and the thermodynamic effects inducing thermal stresses. Doubly rotated quartz resonators present the advantage of one more degree of freedom compared to a singly rotated resonator. The amplitude-frequency behavior is studied as a function of this supplementary parameter which is the rotation angle around the optical axis.

Results are presented for crystals in the range about 5 MHz vibrating on the 3rd and 5th overtone. These resonators are made using plano-convex crystal plates and the influence of the curvature radius is studied. Thus it is possible to define crystallographic cuts and crystal parameters with a lower amplitude-frequency effect.

Introduction

The non linear amplitude-frequency (A-F) effect of quartz crystal resonators is a limiting factor for their use at high drive level in oscillators and filters.

In oscillators at short term the amplitude modulation noise is transformed into frequency noise and for very hard driving the resonator phase response cannot be well defined because of discontinuities. For regular AT cut 5 MHz fifth overtone resonators it is not possible to increase the excitation power above 50 μ W. At long term the all drifts of the driving power are also converted into frequency drifts and, for the same kind of resonators, in the case of power variations of 1 %, to achieve a theoretical relative frequency stability of 1×10^{-12} the mean driving power has to be almost as low as a few nW. Therefore it is not possible to obtain the best short term and long term stabilities at the same time with a single AT cut resonator in the present state of the art.

In hard driven crystal filters, both discrete element and monolithic, intermodulation will occur also due to elastic non linearities and thermal effects. (Sometimes an other phenomenon is present related with surface defects).

In order to cancel out these phenomena several ways are possible. The dual crystal oscillator concept is a promissive possibility which enables to associate with a good spectral purity a low long term drift. But in fact such a "sophisticated" oscillator will be more complicated and then more expensive. A solution is offered by doubly rotated crystals¹, presenting one additional degree of freedom, the rotation angle around the optical axis, which can be chosen in order to decrease the non linear elastic behavior, the other cut angle being adjusted to minimize the thermal sensitivity.

Doubly rotated cut characteristic angles are presented on figure 1. These cuts are characterized by a first rotation φ around the x_3 axis, followed by a second rotation θ around the x'_1 axis.

Amplitude-Frequency effect mechanisms

The A-F effect of quartz crystal resonators was measured by Seed² on AT and BT cut 5 MHz resonators, and by Hammond et al³ on high Q one MHz resonators. The theoretical analysis was performed by Seed² who gave a relation between frequency and current, by Gagnepain⁴, who established the non linear equation of the current through the resonator in the case of the pure thickness model. Then more realistic solutions were given by Tiersten⁵ taking into account the functional dependence along the plate and by Barcus⁶ who solved the wave propagation equation keeping a non linear term previously neglected in the other studies and which enables to describe the frequency variations for low drive levels as well as for high levels.

In these studies are introduced the elastic non linearities by means of the higher order elastic coefficients. In fact it is necessary to go up to the fourth order, because the third order coefficients are null for instance in the case of rotated Y cuts. The non linear equations (wave equation, strains,...) are limited up to the third order. The corresponding solutions present a fast response time.

However the thermal behavior related with a variation of the drive level can also introduce a frequency change which appears as A-F effect. But it is necessary to distinguish between quasistatic temperature behavior and thermodynamic effects.

In the first case, the mean temperature changes and therefore the frequency variations follow the law given by the well known temperature coefficients. This is not a non linear mechanism. It can be reduced by choosing the appropriate cut and using the crystal for instance at a temperature corresponding to the turn over point temperature.

The second thermal phenomenon is the thermodynamic effect. The elastic vibrations induce heat sources within the crystal because of the internal losses. These sources are located at the maximum of the mechanical vibrations. A thermal diffusion takes place and the crystal goes to a new thermal equilibrium⁷ with an inhomogeneous temperature distribution.

Therefore thermal stresses appear and there is a coupling with the propagating wave, due to the third order non linear elastic coefficients, inducing a frequency variation. Strictly speaking it is also necessary to take into account the changes in the crystal size and specific mass. The heat brought by the internal thermal sources diffuses with a time constant τ

$$\tau = \frac{h^2 \rho C_s}{\pi^2 \lambda_{22}} \quad (1)$$

where h is the crystal thickness, ρ the specific mass, C_s the specific heat at constant strain and λ_{22} the diffusion constant. The constant given by the relation (1) corresponds to the case where the crystal surface is in contact with a large heat reservoir at uniform and constant temperature. The value of τ is of the order of 0.3 s. ($h = 1.5$ mm).

In the measurement of the A-F law a third effect is to be considered which is the linear response of the resonator to any amplitude variation. The energy trapped in the resonator cannot change instantaneously and there is a time constant related with the Q factor.

Corresponding to the schematic diagram of fig. 2 the current $i(t)$ through the crystal is given by equation (2) :

$$\frac{d^2 i}{dt^2} + \omega_0^2 i = -\frac{\omega_e}{L_1} \sin \omega t - \left(\frac{R_1 + R}{L_1}\right) \frac{di}{dt} \quad (2)$$

where ω_0 is the crystal angular resonance frequency and ω the driving source angular frequency.

The transient rate relaxation constant is large compared to the oscillation period for resonators with high Q factors and a solution can be chosen with the form given by relation (3) :

$$i = I(t) \cos [\omega t + \phi(t)] = I(t) \cos [\psi(t)] \quad (3)$$

where $I(t)$ and $\phi(t)$, respectively the current amplitude and the phase difference between the current and the driving signal $e = E_0 \cos \omega t$ are slow variable functions of time. A supplementary relation is introduced :

$$\frac{dI}{dt} \cos \psi - I \frac{d\phi}{dt} \sin \psi = 0 \quad (4)$$

The non linear behavior of the resonator is included in the motional parameters L_1 , C_1 and R_1 .

$$\begin{aligned} L_1 &= L_{10} (1 + \alpha_1 I^2 + \dots) \\ C_1 &= C_{10} (1 + \beta_1 I^2 + \dots) \\ R_1 &= R_{10} (1 + \gamma_1 I^2 + \dots) \end{aligned} \quad (5)$$

The parallel capacitor variations are negligible. Using relations (5), (4) and (3) into (2) and taking the average over a period one obtains the mean amplitude \bar{I} and phase $\bar{\phi}$:

$$\frac{d\bar{I}}{dt} = \frac{e_0}{2L_{10}} \cos \bar{\phi} - \frac{R_{10} + R}{2L_{10}} \bar{I} \quad (6)$$

$$\bar{I} \frac{d\bar{\phi}}{dt} = \epsilon \omega \bar{I} - \frac{e_0}{2L_{10}} \sin \bar{\phi} - k \bar{I}^3$$

with $\omega_0 = \omega(1 + \epsilon)$. k is a phenomenological coefficient characteristic of the A-F effect including α_1 , β_1 , γ_1 .

The steady state is given by $d\bar{I}/dt = 0$ and if the resonator is driven by a signal with a frequency such as $\bar{\phi} = 0$ equation (6) becomes :

$$\frac{\Delta f}{f_0} = \epsilon = k I^2 \quad (7)$$

which can be used to determine the value of k .

Equations (6) are solved using a computer. The phase response $\bar{\phi}(t)$ is represented on figure 3a for a 5.6 MHz AT cut fifth overtone resonator with a Q factor equal to 0.85×10^6 . The other parameter values are $L_{10} = 4$ H, $R_{10} = 165 \Omega$, $k = 0.25 \text{ A}^{-2}$ and $R = 50 \Omega$.

Each curve corresponds to the transient rate following a step of the driving signal represented by the averaged current of the steady state.

It can be seen on the experimental curves of figure 3b that for this AT cut resonator the thermal effect does not appear. But for some other resonators thermal effects can become very important and give a phase rotation with a large time constant as shown on the experimental curves of figure 4. These curves correspond to a doubly rotated resonator with cut angles not adjusted at the right values.

The sensitivity to internal heat sources was studied applying a laser beam at different points on the surface of a 10 MHz AT cut resonator⁸ and measuring the corresponding frequency variations. Figure 5 shows response curves for the beam applied at several points along the x_3 axis. The most sensitive parts are near the fixation points and this is due to reaction forces of the fixations. At this point the sensitivity is about 5 ppm/°C. At the plate center it becomes lower, approximately 1 ppm/°C.

Amplitude frequency effect versus crystal orientation

The frequency variation as a function of the driving current through the resonator was measured using the measurement set presented on figure 6.

A frequency synthesizer with a local interpolation oscillator drives the crystal which is in a transmission circuit. The synthesizer can be locked at the resonator resonance frequency using a phase loop. An attenuator enables to drive the crystal at different levels, and measuring for each of these levels the corresponding frequency, the A-F curve can be obtained directly. On the other hand by sweeping the source frequency with a low frequency generator the resonator amplitude and phase resonance curves are plotted.

The samples under test were either crystal blanks put between two auxiliary electrodes with an air gap, or resonators plated and mounted inside a metal enclosure. Each blank was laped to give a plano-convex

shape with a given curvature radius between 500 mm and 150 mm. In the case of the air gap the same curvature was given to the upper electrode.

The measurements were performed as a function of the rotation angle from 0° to 25° . The corresponding angle α was adjusted in order to minimize the thermal effects. The resonance frequencies were in the range between 3 MHz and 6 MHz.

On figure 7 are shown the variation of the Amplitude-Frequency effect of thickness shear 3rd overtone resonators as a function of φ for two values of the curvature radius $r = 200$ mm and $r = 300$ mm. The A-F effect which is given on these curves is defined as the relative difference δ between the resonance frequency value measured for a driving current I of 2 mA and the value obtained by extrapolation for I going to zero. The A-F effect is decreasing and a lower value is obtained for φ near 24° and 25° .

Figure 8 shows the same results as a function of the curvature radius r for $\varphi = 18^\circ, 19^\circ, 20^\circ$ and 21° . For the small radius ($r = 150$ mm) the A-F effects are still function of the crystal orientation, but for $r \geq 400$ mm the effect is becoming very low and keeps about the same value for the four orientations.

On figure 9 are presented the results of measurements corresponding to resonators vibrating on the 5th overtone. In this case the experimental data are more dispersed than for the 3rd overtone, and are not so reproducible, mainly when using the air gap. Nevertheless for curvature radius between 300 and 500 mm all the points are located inside a strip (represented on the figure) which is following the same law as for the previous curves. For $r = 200$ and 150 mm the frequency change is more important but defining a precise law would need more experimental data points.

Figure 10 also shows the A-F effect as a function of the curvature radius for $\varphi = 18^\circ, 19^\circ, 20^\circ$ and 21° . The general variation law is almost the same one.

These results depict the general behavior of the A-F effect and show the influence of the crystal orientation. However the curvature radius of plano-convex plates strongly contribute to the behavior by building up the trapping of the energy at the plate center. To compare the A-F behavior of crystals with different orientations it is necessary to use all the time the same level of driving power. A value of current of 2 mA was used as reference. This will correspond to an excitation at constant power if the resonators keep, whatever orientation is, the same motional elements. In fact these elements change, and particularly the motional capacitor and the motional inductor. Figure 11 shows the variation of the inductor L_1 as a function of φ . This curve is deduced from the results presented by Ballato¹ and depicts the increase of L_1 for a doubly rotated cut at constant frequency with respect to the corresponding inductor of a AT cut. It becomes necessary to take this variation into account if the resonator is not driven exactly at its resonance frequency.

If the results are really compared at constant excitation power, a decrease of the curvature radius induces an energy density increase in the active part of the crystal. The energy per unit volume getting more important the A-F effect is thus increased.

This A-F effect follows a quadratic law as a function of current, which is given by equation (7). The characteristic coefficient k is calculated from the locus of the resonance curve maximum.

Figure 12 shows these amplitude resonance curves for resonators vibrating on their 3rd overtone in the case of AT cut and cuts with $\varphi = 19^\circ$ and $\varphi = 24^\circ$, with curvature radius equal to 300 mm. Figures 13 and 14 present the same resonance curves in the case of the 5th overtone for AT cut and $\varphi = 10^\circ, \varphi = 20^\circ$ and $\varphi = 24^\circ$, using the same curvature radius of 300 mm. The correspond values of the A-F coefficient are given in table 1.

curvature radius $r = 300$ mm	A-F coefficient k (mA^{-2})	crystallographic cut
3rd overtone	0.2×10^{-6}	AT
	0.045×10^{-6}	$\varphi = 19^\circ$
	0.02×10^{-6}	$\varphi = 24^\circ$
5th overtone	0.2×10^{-6}	AT
	0.11×10^{-6}	$\varphi = 10^\circ$
	0.05×10^{-6}	$\varphi = 20^\circ$
	0.02×10^{-6}	$\varphi = 24^\circ$

table 1

Figure 15 shows the resonance curves of a resonator, with a curvature radius equal to 400 mm and $\varphi = 20^\circ$, which has for the 3rd overtone an A-F coefficient k equal to $0.004 \times 10^{-6} \text{ mA}^{-2}$ and for the 5th overtone a coefficient of $0.02 \times 10^{-6} \text{ mA}^{-2}$. These two last results are those presenting the lowest A-F effect among the resonators which were studied. For the 3rd overtone k is 50 times lower as for the corresponding AT cut.

Conclusion

These measurements show it is possible, by the choice of appropriated doubly rotated cut and vibration mode, to decrease by an important factor the A-F effect. However the curvature radius of the plano-convex plate by energy trapping has a large influence on the value of the A-F coefficient k and therefore it is necessary to use a large curvature radius to minimize the non linearities. But trapped energy resonators also need a small curvature radius to obtain high Q factors. It is therefore necessary to choose a compromise between low A-F behavior and high Q factor.

Resonators with low non linear effects will enable to make oscillators with better spectral purity, by increasing the oscillation level and therefore the signal over noise ratio. But also these oscillators will have a better long term stability by decreasing the influence of slow variations of the driving power.

Bibliography

- (1) A. Ballato, G.J. Iafrate, "The Angular Dependence of Piezoelectric Plate Frequencies and their Temperature Coefficients", 30th Ann. Freq. Cont. Symp., Fort Monmouth (1976).

- (2) A. Seed, Fourth International Congress on Acoustics, Copenhagen, Aug (1962).
- (3) D. Hammond, C. Adams, L. Cutler, 17th Ann. Freq. Cont. Symp., Fort Monmouth (1963).
- (4) J.J. Gagnepain, R. Besson, "Non linear effects in piezoelectric quartz crystals", Physical Acoustics, vol. XI, W.P. Mason, R.N. Thurston, Editors, Academic Press, N.Y. (1975).
- (5) H.F. Tiersten, "Analysis of non linear resonance in rotated Y-Cut quartz thickness-shear resonators", 29th Ann. Freq. Cont. Symp. Fort Monmouth (1975).
- (6) L.C. Barcus, "Non linear effects in the AT-cut quartz resonator", IEEE Trans. on Sonics and Ultrasonics, vol. SU 22, n° 4 (1975).
- (7) G. Théobald, J.J. Gagnepain, "Frequency variations due to mechanical heating in quartz crystal resonators", to be published.
- (8) J.J. Gagnepain, "Fundamental noise studies of quartz crystal resonators", 30th Ann. Freq. Cont. Symp., Fort Monmouth (1976).

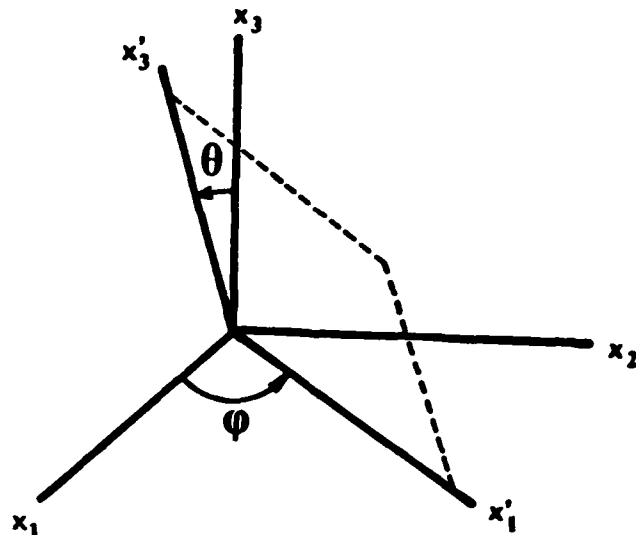


Fig. 1. Angles defining doubly rotated cuts.

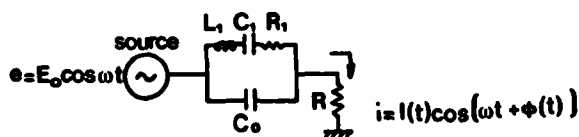


fig. 2

Fig. 2. Schematic for determining crystal current.

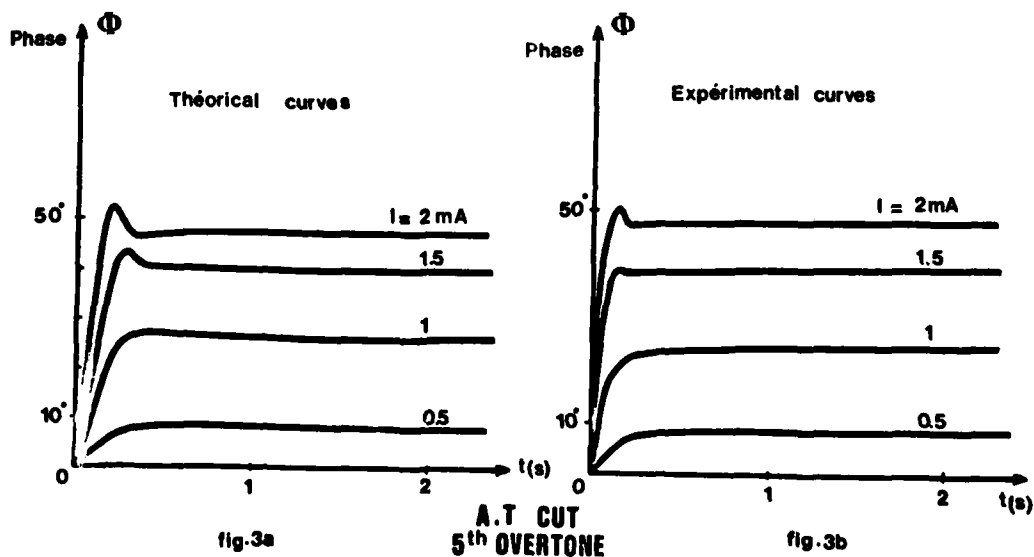


Fig. 3. Theoretical and experimental phase transients following step functions of crystal current.

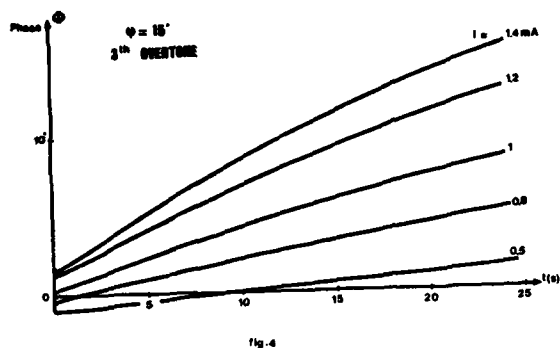


Fig. 4. Experimental, long period phase transients of doubly rotated quartz cut.

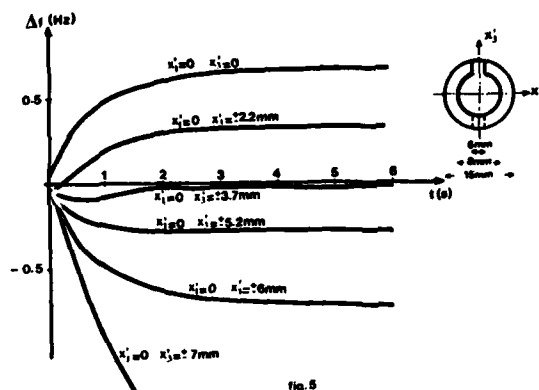


Fig. 5. Frequency transient responses to laser beams applied to different portions of the crystal surface.

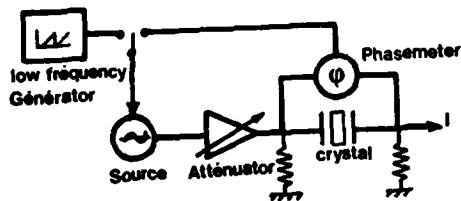


Fig. 6. Measurement setup used to obtain crystal frequency-current response.

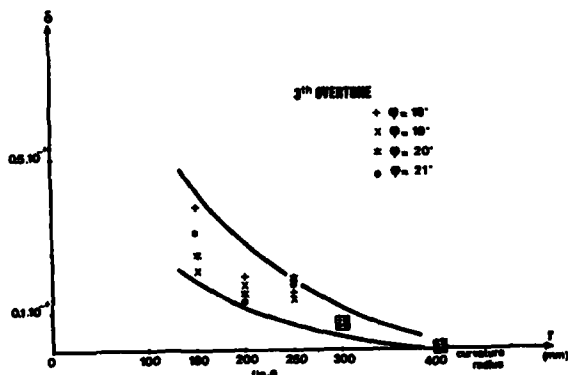


Fig. 8. Variation of amplitude-frequency effect with curvature radius of resonator. (Third overtone resonators.)

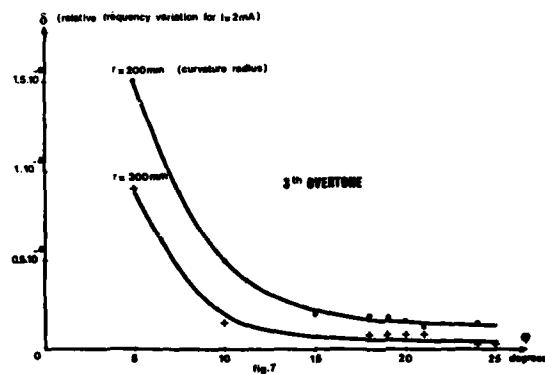


Fig. 7. Variation of amplitude-frequency effect with angle of rotation about optical axis. (Third overtone resonators.)

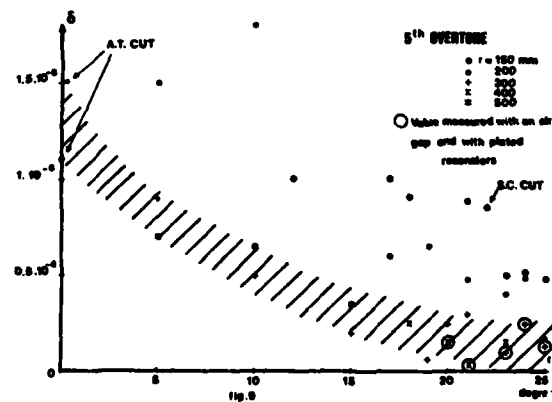


Fig. 9. Variation of amplitude-frequency effect with angle of rotation about optical axis. (Fifth overtone resonators.)

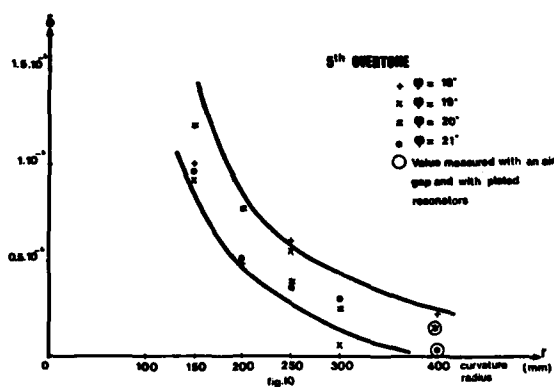


Fig. 10. Variation of amplitude-frequency effect with curvature radius of resonator. (Fifth overtone resonators.)

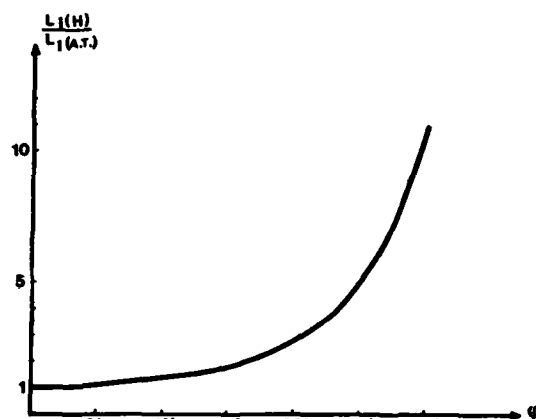


Fig. 11. Motional inductance of doubly rotated cuts normalized to that of an AT-cut having equal frequency

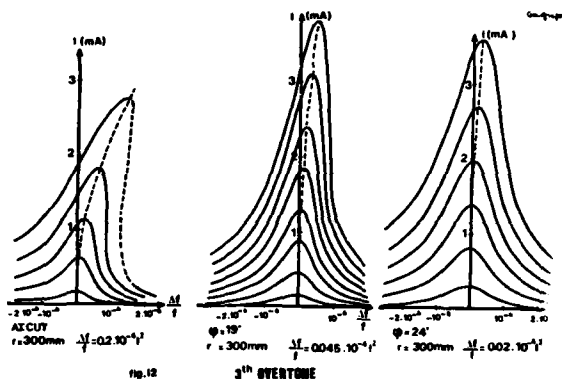


Fig. 12. Amplitude-frequency resonance curves for third overtone units having phi angles of 0° (AT-cut) and 10° .

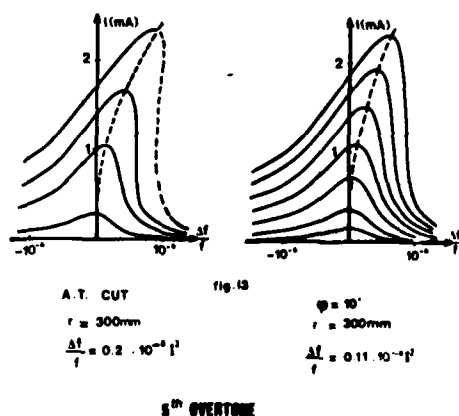


Fig. 13. Amplitude-frequency resonance curves for fifth overtone units having phi angles of 0° (AT-cut) and 10° .

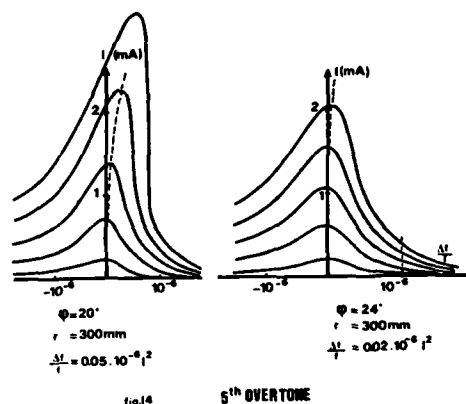


Fig. 14. Amplitude-frequency resonance curves for fifth overtone units having phi angles of 20° and 24° .

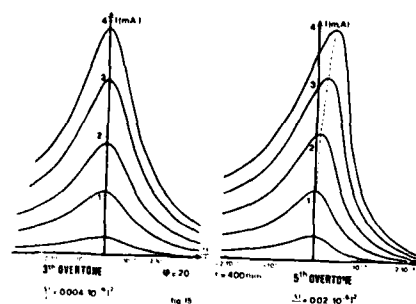


Fig. 15. Amplitude-frequency resonance curves for the third and fifth overtones of a unit having the phi angle of 20° .

TEMPERATURE INDUCED FREQUENCY CHANGES IN ELECTRODED AT-CUT QUARTZ THICKNESS-SHEAR RESONATORS

H.F. Tiersten and B.K. Sinha
Department of Mechanical Engineering,
Aeronautical Engineering & Mechanics
Rensselaer Polytechnic Institute
Troy, New York 12181

Abstract

A perturbation analysis of the general linear electroelastic equations for small dynamic fields superposed on a static bias has been performed. The bias may be either electrical or mechanical or both. The equation for the first perturbation of the eigenvalue obtained in that analysis forms an important part of this analysis. In this work a system of approximate equations for the determination of thermal stresses in piezoelectric plates with thin films of a different material plated on the surfaces is derived. The plate equations are obtained by making a suitable expansion of the pertinent variables in the thickness coordinate, inserting the expansion in the appropriate variational principle and integrating with respect to the thickness in the manner of Mindlin. Conditions resulting in both extensional and flexural stresses are considered. The particular case of purely extensional thermal stresses resulting from electrodes of equal thickness plated on the major surfaces of a rotated Y-cut quartz plate of uniform thickness is treated in detail. The attendant thermal biasing stresses and strains produce a change in resonant frequency when substituted in the aforementioned equation for the first perturbation of the eigenvalue due to a bias. Calculations are performed for gold strip electrodes running in various directions on AT-cut quartz plates.

1. Introduction

A perturbation analysis of the linear electroelastic equations for small fields superposed on a bias has been performed. The change in resonant frequency due to any bias such as, e.g., a residual stress may readily be obtained from the resulting equation for the first perturbation of the eigenvalue if the bias is known. The use of this perturbation equation has already been shown to be extremely accurate in the determination of changes in the surface wave velocity of crystals due to flexural biasing stresses.

In this paper a system of approximate plate equations for the determination of thermal stresses in thin piezoelectric plates coated with much thinner films is derived in the manner of Mindlin³⁻⁵. The resulting approximate equations simplify the treatment of many thermal stress problems considerably, and the three-dimensional detail not included in the approximate description is not deemed to be important for our purposes. Both extensional and flexural equations are obtained. The case of purely extensional thermal stresses arising from identical electrodes on a rotated Y-cut quartz plate is treated in detail and the resulting change in resonant frequency is calculated from the equation for the first perturbation of the eigenvalue. Calculations are performed for gold strip electrodes running in different directions on large flat AT-cut quartz plates.

2. Perturbation Equations

The equation for the first perturbation of the eigenvalue obtained from the perturbation analysis mentioned in the Introduction may be written in the form

$$\Delta \mu = H_\mu / 2\omega_\mu, \quad \omega = \omega_\mu - \Delta \mu, \quad (2.1)$$

where ω_μ and ω are the unperturbed and perturbed eigenfrequencies, respectively, and

$$H_\mu = \int_S N_L [\tilde{K}_{LY}^{\mu} - K_{LY}^{\mu} u_Y + \tilde{D}_L^{\mu} - d_L^{\mu} \tilde{\phi}] ds + \int_V [\tilde{K}_{LY,L}^{\mu} g_Y^{\mu} + \tilde{D}_{L,L}^{\mu}] dv, \quad (2.2)$$

in which S is a surface enclosing a volume V . In (2.2) N_L denotes the unit normal to the undeformed surface at the reference temperature, u_Y denotes the mechanical displacement vector, and $\tilde{\phi}$ denotes the electric potential. The linear stress K_{LY}^{μ} and electric displacement vector D_L^{μ} are given by the usual linear piezoelectric constitutive relations

$$\begin{aligned} \tilde{K}_{LY}^{\mu} &= \hat{c}_{LYM\alpha} u_{\alpha,M} + e_{MYL} \tilde{\phi}, \\ \tilde{D}_L^{\mu} &= e_{LYM} u_{Y,M} - \hat{e}_{LM} \tilde{\phi}, \end{aligned} \quad (2.3)$$

and as usual satisfy

$$\tilde{K}_{LY,L}^{\mu} = \rho \tilde{u}_Y, \quad \tilde{D}_{L,L}^{\mu} = 0. \quad (2.4)$$

The nonlinear contributions to the Piola-Kirchhoff stress tensor \tilde{K}_{LY}^{μ} and reference electric displacement vector \tilde{D}_L^{μ} take the respective forms

$$\begin{aligned} \tilde{K}_{LY}^{\mu} &= \hat{c}_{LYM\alpha} u_{\alpha,M} + \hat{e}_{MYL} \tilde{\phi}, \\ \tilde{D}_L^{\mu} &= \hat{e}_{LYM} u_{Y,M} - \hat{e}_{LM} \tilde{\phi}, \end{aligned} \quad (2.5)$$

where $\hat{c}_{LYM\alpha}$, \hat{e}_{MYL} and \hat{e}_{LM} are effective material constants that depend on the biasing state. Clearly, the total dynamic Piola-Kirchhoff stress tensor and reference electric displacement vector are given by

$$\tilde{K}_{LY}^{\mu} = \tilde{K}_{LY}^{\mu} + \tilde{K}_{LY}^{\mu}, \quad \tilde{D}_L^{\mu} = \tilde{D}_L^{\mu} + \tilde{D}_L^{\mu}. \quad (2.6)$$

The vector g_Y^{μ} denotes the normalized mechanical displacement for the μ th unperturbed mode and $\tilde{\phi}^{\mu}$ denotes the normalized electric potential for the μ th mode, i.e.,

$$g_Y^{\mu} = \frac{u_Y^{\mu}}{N_\mu}, \quad \tilde{\phi}^{\mu} = \frac{\tilde{\phi}^{\mu}}{N_\mu}, \quad (2.7)$$

where

$$N_\mu^2 = \int_V \rho u_Y^{\mu} u_Y^{\mu} dv. \quad (2.8)$$

The normalized stress tensor K_{LY}^{μ} and electric displacement vector d_L^{μ} , both for the normalized μ th mode, are given by the linear piezoelectric constitutive relations (2.3) when u_Y and $\tilde{\phi}$ are replaced by g_Y^{μ} and $\tilde{\phi}^{\mu}$, respectively. The upper cycle notation for many dynamic variables and the capital latin and lower case greek index notation is being employed for consistency with Ref. 1, as is the remainder of the notation in this section.

When the electrical and electroelastic nonlinearities are ignored we have $\hat{e}_{LY} = 0$, $\hat{e}_{LY} = 0$, and

$$\hat{c}_{2LYM\alpha}^1 = T_{LM}^1 \delta_{Y\alpha} + c_{2LYM\alpha B}^1 E_{AB}^1 + c_{2LYM\alpha}^1 \alpha_K + c_{2LYM\alpha}^1 w_{Y,K}, \quad (2.9)$$

where

$$T_{LM}^1 = c_{2LMKN}^1 E_{KN}^1, \quad E_{KN}^1 = \frac{1}{2} (w_{K,N} + w_{N,K}) \quad (2.10)$$

and w_K denotes the static biasing displacement field. Thus, in this description the present position χ is related to the reference position $\bar{\chi}$ by

$$\chi(X_L, t) = \bar{\chi} + \bar{w}(X_L) + \bar{u}(X_L, t). \quad (2.11)$$

It has been shown⁶ that for the cases of interest here, (2.2) may be replaced by the considerably simpler form

$$H_\mu = - \int_V \hat{c}_{LYM\alpha}^n g_{Y,L}^\mu dv = - \int_V \hat{c}_{2LYM\alpha}^n g_{Y,L}^\mu dv. \quad (2.12)$$

3. Static Plate Equations

A schematic diagram of the plated crystal plate is shown in Fig. 1 along with the associated coordinate system. Referred to this coordinate system the static form of Mindlin's plate equations may be written³

$$T_{AJ,A}^{(n)} - nT_{2J}^{(n-1)} + F_J^{(n)} = 0, \quad n=0,1,2, \quad (3.1)$$

where A, B, C, D take the values 1 and 3 and skip 2 and

$$T_{IJ}^{(n)} = \int_{-h}^h x_2^n T_{IJ}^{(n)} dx_2, \quad F_J^{(n)} = [x_2^n T_{2J}^{(n)}]_{-h}^h. \quad (3.2)$$

The linear thermoelastic constitutive equations for the n th order stress resultants for homogeneous temperature excursions θ take the form

$$T_{IJ}^{(n)} = c_{IJKL} \sum_{n=0}^2 H_{mn} S_{KL}^{(n)} - d_{IJ}^{(m)} v_{IJ} \theta, \quad (3.3)$$

where we have taken the liberty of dropping the lower script 2 on the c_{IJKL} , the v_{IJ} denote the thermoelastic constants and

$$H_{mn} = \begin{cases} 2h^{(m+n+1)}/(m+n+1) & m+n \text{ even} \\ 0 & m+n \text{ odd} \end{cases}, \quad (3.4)$$

$$d_{IJ}^{(m)} = \begin{cases} 2h^{(m+1)}/(m+1) & m \text{ even} \\ 0 & m \text{ odd} \end{cases}. \quad (3.5)$$

The n th order plate strains take the usual form

$$S_{KL}^{(n)} = \frac{1}{2} [w_{K,L}^{(n)} + w_{L,K}^{(n)} + (n+1)(\delta_{2L} w_K^{(n+1)} + \delta_{2K} w_L^{(n+1)})], \quad (3.6)$$

where

$$w_K = \sum_{n=0}^2 x_2^n w_K^{(n)}. \quad (3.7)$$

The thermoelastic constants v_{IJ} are related to the coefficients of linear expansion α_{KL} by the usual relation⁷

$$v_{IJ} = c_{IJKL} \alpha_{KL}. \quad (3.8)$$

The static form of Mindlin's simplified extensional equations for the very thin electrode platings may be written⁸

$$\delta_{JB} T_{AB,A}^{(0)} + F_J^{(0)} = 0, \quad F_J^{(0)} = [T_{2J}^{(0)}]_{-h}^h, \quad (3.9)$$

where

$$T_{AB}^{(0)} = 2h' \gamma'_{ABCD} S_{CD}^{(0)} - 2h' v_{AB}^{*'} \theta, \quad (3.10)$$

and since the platings are isotropic in the plane

$$\gamma'_{ABCD} = \lambda' \delta_{AB} \delta_{CD} + \mu' (\delta_{AC} \delta_{BD} + \delta_{AD} \delta_{BC}),$$

$$\lambda'_0 = 2\mu' \lambda' / (\lambda' + 2\mu'), \quad v_{AB}^{*'} = [2\mu' (3\lambda' + 4\mu') / (\lambda' + 2\mu')] \alpha'_{AB}, \quad (3.11)$$

which are for the upper plating and similar equations hold for the bottom plating, but with double primes replacing the primes. In order to obtain the static equations for the plated crystal plate we need the boundary conditions at all interfaces. The traction conditions take the form

$$T_{2J}^{(0)}(h') = 0, \quad T_{2J}^{(0)}(-h'') = 0, \\ T_{2J}^{(0)}(-h') = T_{2J}^{(0)}(h), \quad T_{2J}^{(0)}(h'') = T_{2J}^{(0)}(-h), \quad (3.12)$$

and the conditions of continuity of mechanical displacement take the form

$$w_I^{(0)} = w_I^{(0)} + h w_I^{(1)} + h^2 w_I^{(2)}, \\ w_I^{(0)} = w_I^{(0)} - h w_I^{(1)} + h^2 w_I^{(2)}. \quad (3.13)$$

At this point in order to simplify the treatment somewhat, we ignore the elastic constants that cause coupling between shear and extension in the constitutive equations⁹. In obtaining the extensional equations we first allow for the free thickness strain by making the assumptions¹⁰

$$T_{22}^{(0)} = 0, \quad T_{2A}^{(1)} = 0, \quad T_{22}^{(2)} = 0, \quad (3.14)$$

the second of which is required for consistency by virtue of (3.1). Employing (3.14) in the reduced form of (3.3) in the usual manner³, we obtain the extensional constitutive equations in the form

$$T_{11}^{(0)} = 2h[c_{11}^* S_{11}^{(0)} + c_{13}^* S_{33}^{(0)}] + \frac{2}{3} h^3 [c_{11}^* S_{11}^{(2)} + c_{13}^* S_{33}^{(2)}] - 2h v_{11}^* \theta, \\ T_{33}^{(0)} = 2h[c_{13}^* S_{11}^{(0)} + c_{33}^* S_{33}^{(0)}] + \frac{2}{3} h^3 [c_{13}^* S_{11}^{(2)} + c_{33}^* S_{33}^{(2)}] - 2h v_{33}^* \theta, \\ T_{11}^{(2)} = \frac{2}{3} h^3 [c_{11}^* S_{11}^{(0)} + c_{13}^* S_{33}^{(0)}] + \frac{2}{5} h^5 [c_{11}^* S_{11}^{(2)} + c_{33}^* S_{33}^{(2)}] - \frac{2}{3} h^3 v_{11}^* \theta, \\ T_{33}^{(2)} = \frac{2}{3} h^3 [c_{13}^* S_{11}^{(0)} + c_{33}^* S_{33}^{(0)}] + \frac{2}{5} h^5 [c_{13}^* S_{11}^{(2)} + c_{33}^* S_{33}^{(2)}] - \frac{2}{3} h^3 v_{33}^* \theta, \\ T_{13}^{(0)} = 2h c_{55} [S_{13}^{(0)} + \frac{h^2}{3} S_{13}^{(2)}], \\ T_{13}^{(2)} = 2h^3 c_{55} [\frac{1}{3} S_{13}^{(0)} + \frac{h^2}{5} S_{13}^{(2)}], \quad (3.15)$$

where¹¹

$$c_{11}^* = c_{11} - \frac{c_{12}^2}{c_{22}}, \quad c_{13}^* = c_{13} - \frac{c_{12} c_{23}}{c_{22}}, \quad c_{33}^* = c_{33} - \frac{c_{23}^2}{c_{22}}, \\ v_{11}^* = v_{11} - v_{22} c_{12} / c_{22}, \quad v_{33}^* = v_{33} - v_{22} c_{23} / c_{22}. \quad (3.16)$$

From (3.1), with (3.14)₁₋₂, we obtain the second order extensional equations¹⁰

$$T_{AB,A}^{(0)} + F_B^{(0)} = 0, \quad T_{AB,A}^{(2)} + F_B^{(2)} = 0. \quad (3.17)$$

In this approximation from (3.6) the pertinent plate strains $S_{ij}^{(n)}$ ($n=0,2$) take the form

$$S_{11}^{(n)} = w_{1,1}^{(n)}, \quad S_{33}^{(n)} = w_{3,3}^{(n)}, \quad S_{13}^{(n)} = \frac{1}{2} (w_{1,3}^{(n)} + w_{3,1}^{(n)}). \quad (3.18)$$

When written out the constitutive equations, (3.10) for the upper electrode plating take the form

$$\begin{aligned} T_{11}^{(0)'} &= 2h'[(\lambda_o' + 2\mu')w_{1,1}^{(0)'} + \lambda_o'w_{3,3}^{(0)'}] - 2h'\nu'\theta, \\ T_{33}^{(0)'} &= 2h'[(\lambda_o' + 2\mu')w_{3,3}^{(0)'} + \lambda_o'w_{1,1}^{(0)'}] - 2h'\nu'\theta, \\ T_{13}^{(0)'} &= 2h'\mu'(w_{1,3}^{(0)'} + w_{3,1}^{(0)'}), \end{aligned} \quad (3.19)$$

and similar equations exist for the lower electrode plating, but with primes replaced by double primes. In the case of these extensional equations the displacements $w_A^{(0)'}$ and $w_A^{(0)''}$, in the upper and lower electrode platings, respectively, are given by

$$w_A^{(0)'} = w_A^{(0)} + h^2 w_A^{(2)}, \quad w_A^{(0)''} = w_A^{(0)} + h^2 w_A^{(2)}. \quad (3.20)$$

The equations for the plated crystal plate are obtained by employing (3.12) in (3.9) for both the upper and lower platings and then inserting (3.9) for both platings in (3.2)₂, with the result

$$F_B^{(0)} = T_{AB,A}^{(0)'} + T_{AB,A}^{(0)''}, \quad F_B^{(2)} = h^2 (T_{AB,A}^{(0)'} + T_{AB,A}^{(0)''}). \quad (3.21)$$

The substitution of (3.21) into (3.17) yields

$$[T_{AB}^{(0)} + T_{AB}^{(0)'} + T_{AB}^{(0)''}]_{,A} = 0, \quad [T_{AB}^{(2)} + h^2 (T_{AB}^{(0)'} + T_{AB}^{(0)''})]_{,A} = 0, \quad (3.22)$$

which are the second order extensional equations of equilibrium of the plated crystal plate. Clearly on an edge of the plated crystal plate, the traction boundary conditions that accompany (3.22) are¹⁰

$$\begin{aligned} N_A [T_{AB}^{(0)} + T_{AB}^{(0)'} + T_{AB}^{(0)''}] &= \bar{T}_B^{(0)}, \\ N_A [T_{AB}^{(2)} + h^2 (T_{AB}^{(0)'} + T_{AB}^{(0)''})] &= \bar{T}_B^{(2)}, \end{aligned} \quad (3.23)$$

where the $\bar{T}_B^{(n)}$ represent applied traction resultants on the edge. The alternative displacement conditions are on $w_B^{(0)}$ and $w_B^{(2)}$.

Since the upper and lower electrodes can have different thicknesses $2h'$ and $2h''$, flexure can occur even though the temperature change θ is homogeneous. In obtaining the flexural equations we first allow for the free thickness-strain by making the assumptions¹⁰

$$T_{22}^{(1)} = 0, \quad T_{2A}^{(2)} = 0, \quad (3.24)$$

the second of which is required for consistency by virtue of (3.1). In addition, as usual in the elementary theory of flexure we take the thickness-shear plate strains $S_{2A}^{(0)}$ to vanish¹², which with (3.6) yields

$$w_A^{(1)} = -w_{2,A}^{(0)}. \quad (3.25)$$

Employing (3.24) in the reduced form of (3.3) in the usual manner³, we obtain the flexural constitutive

equations in the form

$$\begin{aligned} T_{11}^{(1)} &= \frac{2}{3} h^3 [c_{11}^* S_{11}^{(1)} + c_{13}^* S_{33}^{(1)}], \quad T_{13}^{(1)} = \frac{2}{3} h^3 c_{55}^* S_{13}^{(1)} \\ T_{33}^{(1)} &= \frac{2}{3} h^3 [c_{13}^* S_{11}^{(1)} + c_{33}^* S_{33}^{(1)}], \end{aligned} \quad (3.26)$$

where c_{11}^* , c_{33}^* and c_{13}^* are given in (3.16). From (3.6) and (3.25) we find that the pertinent plate strains take the form

$$S_{AB}^{(1)} = \frac{1}{2} (w_{A,B}^{(1)} + w_{B,A}^{(1)}) = -w_{2,AB}^{(0)}. \quad (3.27)$$

From (3.1), with (3.24)₁, we obtain the usual flexural equations

$$T_{B2,B}^{(0)} + F_2^{(0)} = 0, \quad T_{AB,A}^{(1)} - T_{2B}^{(0)} + F_B^{(1)} = 0, \quad (3.28)$$

where we have taken proper account of the fact that the thickness-shear stress resultants $T_{2A}^{(0)}$ exist even though the associated thickness-shear plate strains $S_{2A}^{(0)}$ vanish. The substitution of (3.28)₂ into (3.28)₁ yields

$$T_{AB,AB}^{(1)} + F_{B,B}^{(1)} + F_2^{(0)} = 0, \quad (3.29)$$

which is the equation of the flexure of thin plates. The equation of flexure for the plated crystal plate is obtained by employing (3.12) in (3.9) for both the upper and lower platings and then inserting (3.9) for both platings in (3.2)₂, with the result

$$F_2^{(0)} = 0, \quad F_B^{(1)} = h [T_{AB,A}^{(0)'} - T_{AB,A}^{(0)''}]. \quad (3.30)$$

In this case of flexure the displacements $w_A^{(0)'}$ and $w_A^{(0)''}$, in the upper and lower electrode platings, respectively, are given by

$$w_B^{(0)'} = h w_B^{(1)} = -h w_{2,B}^{(0)}, \quad w_B^{(0)''} = -h w_B^{(1)} = h w_{2,B}^{(0)}. \quad (3.31)$$

The substitution of (3.30) into (3.29) yields

$$[T_{AB}^{(1)} + h (T_{AB}^{(0)'} - T_{AB}^{(0)''})]_{,AB} = 0, \quad (3.32)$$

which is the equation of static flexure for the plated crystal plate. The edge conditions for the plated plate are the same as in the elementary theory of flexure of thin plates when the moments are given by the quantities in brackets in (3.32).

4. Strip Electrodes on AT-Cut Quartz Plates

In this section we apply the equations presented in the last two sections to the case of identical strip electrodes running in an arbitrary direction on an AT-cut quartz plate. A schematic diagram of the electrode plate is shown in Fig. 2. The \hat{x}_1 -coordinate axis shown in the figure makes an angle γ with the diagonal axis. Since the plate is not constrained on the free edges normal to the \hat{x}_3 -axis, we introduce the conditions^{9,10,13}

$$\hat{T}_{33}^{(0)} = 0, \quad \hat{T}_{33}^{(2)} = 0, \quad \hat{T}_{13}^{(0)} = 0, \quad \hat{T}_{13}^{(2)} = 0. \quad (4.1)$$

which permit free-expansion in the \hat{x}_3 -direction. From (3.15), with (4.1), we obtain the non-trivial constitutive equations

$$\begin{aligned} \hat{T}_{11}^{(0)} &= 2h \hat{c}_{11}^{**} \hat{S}_{11}^{(0)} + \frac{2}{3} h^3 \hat{c}_{11}^{***} \hat{S}_{11}^{(2)} - 2h \hat{c}_{11}^{**} \nu, \\ \hat{T}_{11}^{(2)} &= \frac{2}{3} h^3 \hat{c}_{11}^{***} \hat{S}_{11}^{(0)} + \frac{2}{5} h^5 \hat{c}_{11}^{****} \hat{S}_{11}^{(2)} - \frac{2}{3} h^3 \hat{c}_{11}^{***} \nu, \end{aligned} \quad (4.2)$$

$$\hat{c}_{11}^{**} = \hat{c}_{11}^* - \frac{(\hat{c}_{13}^*)^2}{\hat{c}_{33}^*}, \quad \hat{v}_{11}^{**} = \hat{v}_{11}^* - \frac{\hat{v}_{33}^* \hat{c}_{13}^*}{\hat{c}_{33}^*}. \quad (4.3)$$

Similarly, introducing the conditions

$$\hat{T}_{33}^{(0)'} = 0, \quad \hat{T}_{13}^{(0)'} = 0, \quad (4.4)$$

in the electrode plating, from (3.10) and (3.11) we obtain

$$\hat{T}_{11}^{(0)'} = 2h'E'\hat{S}_{11}^{(0)'} - 2h'\alpha'E'\theta, \quad (4.5)$$

where E' is Young's modulus for the electrode material. Under these circumstances the non-trivial differential equations (3.22) take the form

$$[\hat{T}_{11}^{(0)} + 2\hat{T}_{11}^{(0)'}]_{,1} = 0, \quad [\hat{T}_{11}^{(2)} + 2h'\hat{T}_{11}^{(0)'}]_{,1} = 0, \quad (4.6)$$

in the electroded region. Since all stress resultants vanish in the unelectroded region, from the boundary conditions (3.23) we have

$$\hat{T}_{11}^{(0)} + 2\hat{T}_{11}^{(0)'} = 0, \quad \hat{T}_{11}^{(2)} + 2h'\hat{T}_{11}^{(0)'} = 0 \text{ at } \hat{x}_1 = l, \quad (4.7)$$

and (4.6) is satisfied if (4.7) holds for all $|\hat{x}_1| < l$. Substituting from (4.2) and (3.18) into (4.7), we obtain

$$a_{11}\hat{w}_{1,1}^{(0)} + a_{12}\hat{w}_{1,1}^{(2)} = A^{(0)}\theta, \quad a_{12}\hat{w}_{1,1}^{(0)} + a_{22}\hat{w}_{1,1}^{(2)} = A^{(2)}\theta, \quad (4.8)$$

where

$$\begin{aligned} a_{11} &= 2[h\hat{c}_{11}^{**} + 2h'E'], \quad a_{12} = 2h'\left[\frac{h}{3}\hat{c}_{11}^{**} + 2h'E'\right] \\ a_{22} &= 2h'\left[\frac{h}{15}\hat{c}_{11}^{**} + 2h'E'\right], \quad A^{(0)} = 2(h\hat{v}_{11}^{**} + 2h'\alpha'E') \\ A^{(2)} &= 2h'\left[\frac{h}{3}\hat{v}_{11}^{**} + 2h'\alpha'E'\right]. \end{aligned} \quad (4.9)$$

Equations (4.8) constitute two linear inhomogeneous equations in $\hat{w}_{1,1}^{(0)}$ and $\hat{w}_{1,1}^{(2)}$, which may readily be solved for a given temperature change θ . In order to calculate the change in frequency from this solution of the approximate plate equations, we must have the associated three-dimensional displacement field \hat{w}_k , which to the order of approximation employed takes the form

$$\hat{w}_A = \hat{w}_A^{(0)} + X_2^2 \hat{w}_A^{(2)}, \quad \hat{w}_2 = X_2 \hat{w}_2^{(1)} + X_2^3 \hat{w}_2^{(3)}, \quad (4.10)$$

where $\hat{w}_1^{(0)}$ and $\hat{w}_1^{(2)}$ are determined from the solution of (4.8) and

$$\begin{aligned} \hat{w}_{3,3}^{(0)} &= -\frac{\hat{c}_{13}^*}{\hat{c}_{33}^*} \hat{w}_{1,1}^{(0)} + \frac{\hat{v}_{33}^*}{\hat{c}_{33}^*} \theta, \quad \hat{w}_{3,3}^{(2)} = -\frac{\hat{c}_{13}^*}{\hat{c}_{33}^*} \hat{w}_{1,1}^{(2)}, \\ \hat{w}_2^{(1)} &= -\left[\frac{\hat{c}_{12}^*}{\hat{c}_{22}^*} - \frac{\hat{c}_{23}\hat{c}_{13}^*}{\hat{c}_{22}\hat{c}_{33}^*}\right] \hat{w}_{1,1}^{(0)} + \left[\frac{\hat{v}_{22}^*}{\hat{c}_{22}^*} - \frac{\hat{c}_{23}\hat{v}_{33}^*}{\hat{c}_{22}\hat{c}_{33}^*}\right] \theta, \\ \hat{w}_2^{(3)} &= -\frac{1}{3} \left[\frac{\hat{c}_{21}^*}{\hat{c}_{22}^*} - \frac{\hat{c}_{23}\hat{c}_{13}^*}{\hat{c}_{22}\hat{c}_{33}^*}\right] \hat{w}_{1,1}^{(2)}. \end{aligned} \quad (4.11)$$

The change in eigenfrequency for an AT-cut quartz thickness-shear resonator due to strip electrodes may now be obtained by substituting from (4.10), with (4.11), into the pertinent equations in Sec. 2. When specialized to the case of the n th harmonic of thickness-shear, Eqs. (2.1) may be rewritten in the form

$$\Delta_n = H_n/2\omega_n, \quad \omega = \omega_n - \Delta_n. \quad (4.12)$$

For the n th harmonic thickness-shear mode¹⁴ it is well known that to a very high degree of accuracy¹⁰ we have

$$u_{1n} = \sin \frac{n\pi}{2h} x_2 e^{\frac{i\omega}{n} t}, \quad \omega_n = \frac{n\pi}{2h} \sqrt{\frac{c_{66}}{\rho}}, \quad (4.13)$$

where

$$\bar{c}_{66} = c_{66} + e_{26}^2/c_{22}. \quad (4.14)$$

From (4.13)₁ and (2.8) we have

$$N_n^2 = \rho h, \quad (4.15)$$

which with (2.7)₁ and (4.13)₁ yields

$$g_{1n} = (1/\rho h)^{1/2} [\sin n\pi x_2/2h] e^{\frac{i\omega}{n} t}, \quad n = 1, 3, 5, \dots \quad (4.16)$$

Substituting from (4.10), with (4.11), into (2.9) and (2.10) and then from (2.9) and (4.16) into (2.12) and performing the integration with respect to thickness¹⁰, we obtain

$$\begin{aligned} H_n &= -\frac{1}{\rho n\pi} \left[\left[(2c_{66} + c_{166}) \sin^2 \gamma + c_{366} \cos^2 \gamma \right] \hat{w}_{1,1}^{(0)} \right. \\ &+ c_{266} \hat{w}_2^{(1)} + \left[(2c_{66} + c_{166}) \cos^2 \gamma + c_{366} \sin^2 \gamma \right] \hat{w}_{3,3}^{(0)} \left. \frac{n^3 \pi^3}{4h^2} \right. \\ &+ \left[(2c_{66} + c_{166}) \sin^2 \gamma + c_{366} \cos^2 \gamma \right] \hat{w}_{1,1}^{(2)} + 3c_{266} \hat{w}_2^{(3)} \\ &\left. + \left[(2c_{66} + c_{166}) \cos^2 \gamma + c_{366} \sin^2 \gamma \right] \hat{w}_{3,3}^{(2)} \right] \frac{n\pi}{2} \left(\frac{n^2 \pi^2}{6} - 1 \right). \end{aligned} \quad (4.17)$$

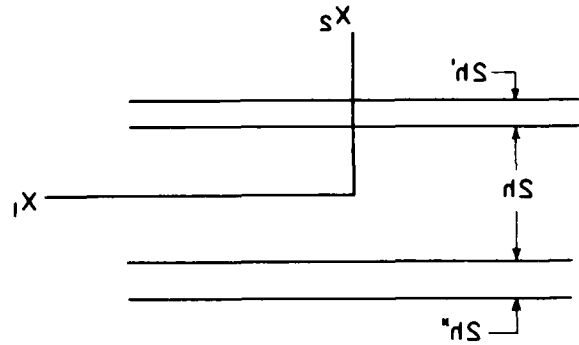
The change in frequency with temperature due to identical strip electrodes of thickness $2h'$ plated on large flat AT-cut quartz plates may now be calculated from (4.17) and (4.12) for strips making an angle γ with respect to the digonal axis. The calculation is restricted to AT-cut quartz plates because the change in the elastic constants with temperature has not been included in the analysis and presumably the change in frequency due to the change in constants with temperature is exactly compensated by the change in frequency due to changes in strain with temperature for the unelectroded AT-cut quartz plate. Consequently, for the AT-cut quartz plate with strip electrodes, the change in frequency with temperature due to identical electrodes of thickness $2h'$ is obtained by subtracting the change in frequency obtained from (4.17) and (4.12) for $h' = 0$ from that obtained using the known value of $2h'$. Since certain anisotropic constants, which we deem to be relatively unimportant, have been ignored in this treatment¹⁰⁻¹¹, there is a small dependence in the change in frequency on the angle γ for the case $h' = 0$, which in reality cannot exist. However, since this same unwarranted anisotropic effect, which is a result of the analytical procedure, exists in the case of the actual electrode thickness $2h'$ also and the two calculations are subtracted to obtain the result, this small unwarranted anisotropy in the calculation is insignificant¹⁰ insofar as the end results are concerned. Calculations have been performed using the known third order elastic constants¹⁵ for the case of 4000 Å thick gold strip electrodes on a fifth overtone 5 MHz AT-cut quartz plate as a function of the angle γ between the strip and the digonal axis and the results are plotted in Fig. 3. The curve has the symmetry shown in the figure because the digonal axis (X_1) is a two-fold rotation axis. In Fig. 4 the calculated change in resonant frequency for gold strip electrodes running normal to the digonal direction ($\gamma = \pi/2$) is plotted as a function of the thickness of the electrodes.

Acknowledgements

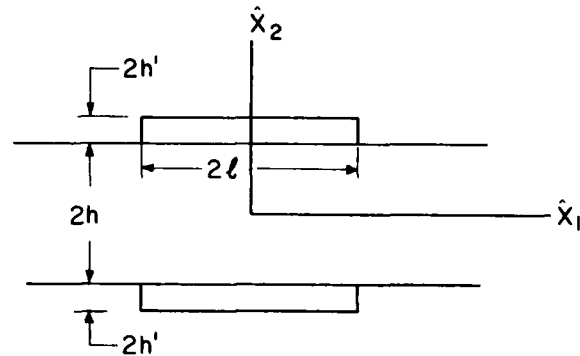
This work was supported in part by the Army Research Office under Grant No. DAAG29-76-G-0173, the Office of Naval Research under Contract No. N00014-76-C-0368 and the National Science Foundation under Grant No. ENG 72-04223.

References

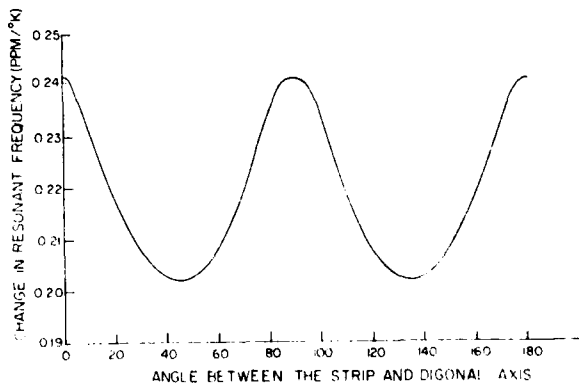
1. H.F. Tiersten, "Perturbation Theory for Linear Electroelastic Equations for Small Fields Superposed on a Bias," to be issued as a technical report, Rensselaer Polytechnic Institute, Troy, N.Y. 12181.
2. B.K. Sinha and H.F. Tiersten, "On the Influence of Uniaxial Biasing Stresses on the Velocity of Piezoelectric Surface Waves," 1976 Ultrasonics Symposium Proceedings, IEEE Cat. No. 76CH112-5SU, Institute of Electrical and Electronics Engineers, New York, 1976, p.475.
3. R.D. Mindlin, "An Introduction to the Mathematical Theory of the Vibration of Elastic Plates," U.S. Army Signal Corps Eng. Lab., Fort Monmouth, NJ (1955). Signal Corps Contract DA-36-039SC-56772.
4. R.D. Mindlin, "High Frequency Vibrations of Crystal Plates," Quart. Appl. Math., 19, 51 (1961).
5. H.F. Tiersten, Linear Piezoelectric Plate Vibrations (Plenum, New York, 1969), Chap.13.
6. H.F. Tiersten and B.K. Sinha, "Temperature Induced Frequency Changes in Electroded AT-Cut Quartz Thickness-Shear Resonators," to be issued as a technical report, Rensselaer Polytechnic Institute, Troy, N.Y. 12181.
7. J. Tasi and G. Herrmann, "Thermoelastic Dissipation in High Frequency Vibrations of Crystal Plates," J. Acoust. Soc. Am., 36, 100 (1964).
8. R.D. Mindlin, "High Frequency Vibrations of Plated Crystal Plates," in Progress in Applied Mechanics (Macmillan, New York, 1963), pp.73-84.
9. This approximation, which we do not believe results in any serious loss in accuracy in most cases, can readily be eliminated⁶, but we do not bother with this refinement here.
10. For more detail see Ref.6.
11. In this treatment we ignore shear thermoelastic constants, in particular v_{13} , which can readily be included⁶, but we do not consider this refinement here. We do not believe that this approximation results in any serious loss in accuracy.
12. R.D. Mindlin, Ref.3, Chap.6.
13. In order to eliminate these assumptions in the electroded region, we must consider both electrode dimensions. This is done in Ref.6 when the shear thermoelastic constants are ignored.
14. For a wide plate with wide electrodes the influence of energy trapping can be ignored.
15. R.N. Thurston, H.J. McSkimin and P. Andreatch, Jr., "Third Order Elastic Constants of Quartz," J. Appl. Phys., 37, 267 (1966).



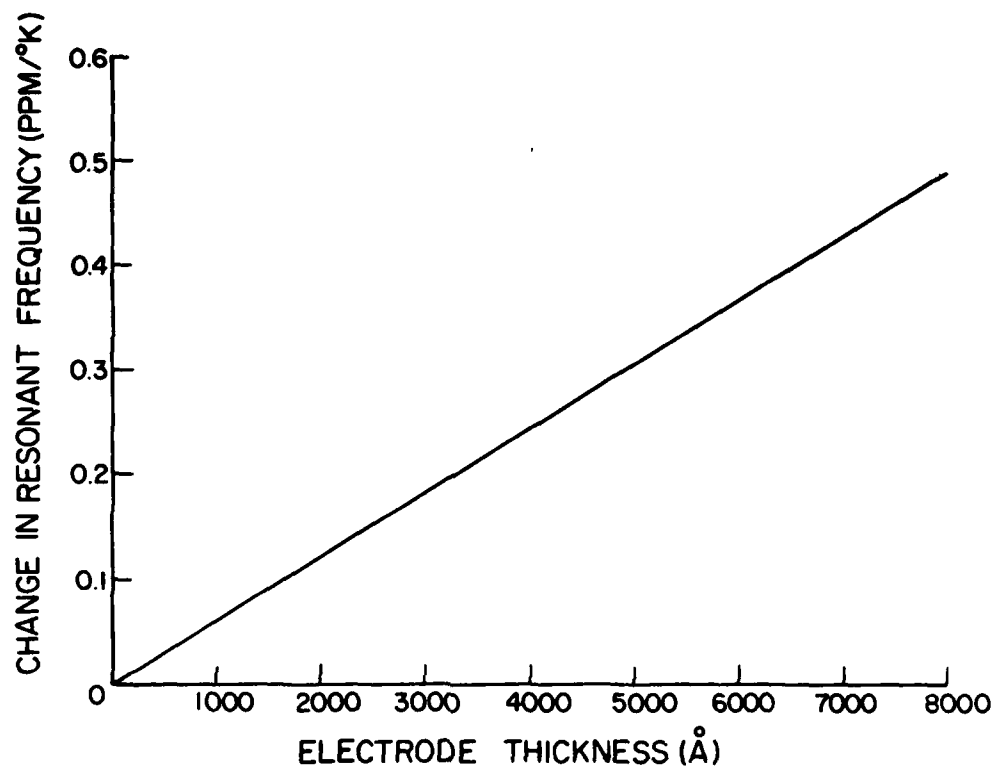
1. Schematic Diagram of the Plated Crystal Plate



2. Schematic Diagram of the Plate Coated with Identical Strip Electrodes



3. Change in Resonant Frequency per °K for Wide Gold Strip Electrodes 4000 Å Thick on a Large Flat AT-Cut Quartz Plate as a Function of the Angle Between the Strip and the Diagonal Axis



4. Change in Resonant Frequency per °K for Wide Gold Strip Electrodes Running Normal to the Digonal Axis on a Large AT-Cut Quartz Plate as a Function of the Electrode Thickness

The Influence of Support-Configuration on the Acceleration Sensitivity of Quartz Resonator Plates

by

P. C. Y. Lee and Kuang-Ming Wu

Department of Civil Engineering
Princeton University
Princeton, N.J. 08540

Summary

The changes in the thickness-shear resonance frequencies of a circular, rotated Y-cut of quartz plate subject to in-plane steady accelerations was found to be a function of the direction and magnitude of the acceleration, the orientation of the crystal cut, the diameter-to-thickness ratio of the plate, and the numbers and positions of the metal ribbon supports at the edge of the crystal.

The purpose of the present study is to investigate how the support-configuration will influence the acceleration sensitivity of a resonator. An AT-cut circular quartz plate with four metal ribbon supports is chosen for a systematic parametric study.

In order to examine the influence of the support-configurations on the frequency changes, the support-configuration is changed from a reference configuration by displacing, or instance, one support in increments of 15° counter-clockwise while holding the other supports fixed. For each displaced configuration, the frequency changes due to acceleration are computed as a function of the acceleration direction. In the same manner, computations are performed for a total of seven series of configuration variations. Then in each series, the particular support-configurations and acceleration directions are identified as giving maximum and zero acceleration effects.

Introduction

In a previous paper¹, the changes in the resonance frequencies of the coupled thickness-shear and flexural vibrations of a circular, rotated Y-cut of quartz subject to in-plane steady accelerations were studied by employing the equations of plate vibrations² in which the initial fields of stress and strain and the nonlinear stress-strain relations contributed by the third-order elastic stiffnesses are taken into account. The changes in the resonance frequencies due to accelerations were obtained by a perturbation method. The predicted values were compared to the experimentally measured ones³ with good agreement.

It was found that the changes in the thickness-shear resonance frequencies due to the initial accelerations of a circular plate supported by metal ribbons around the edge (See Fig. 1) depend upon the direction and magnitude of the acceleration, the orientation of the crystal cut,

the diameter-to-thickness ratio of the plate, and the number of supports and their positions. This functional relation may be expressed as follows.

$$\frac{\Delta\Omega}{\Omega_0} = \frac{\Delta f}{f_0} = F(\psi, G, \theta, R/b, \alpha_i, n) \quad (1)$$

where

- f_0 thickness-shear and flexural resonance frequency of the plate with no acceleration,
- Δf frequency changes due to acceleration,
- $\Omega_0, \Delta\Omega$ dimensionless frequency and its changes,
- ψ angle denoting the direction of the acceleration with respect to the x_1 -axis,
- G magnitude of the body force per unit volume which is equivalent to the product of mass density of the crystal and the acceleration,
- θ orientation of the rotated Y-cut,
- R/b diameter-to-thickness ratio,
- α_i angles denoting the positions of the supports, with respect to the x_1 -axis, and $i = 1, 2, \dots, n$,
- n total number of supports, which can be any positive integer greater than two.

Details of this problem, the analytical formulation and its solution are given in Refer. 1.

Parametric Study

It is seen from (1) that the frequency changes may be affected by a total of eight parameters for a rotated Y-cut plate with four supports. The effect of each one of these parameters may be studied by computing, one by one, the frequency changes vs. a range of values of a chosen parameter while holding the values of the rest seven parameters constant. Although the procedure is straight forward, it shall require a tremendous amount of time in computation, and in presentation and in interpretation of the results.

In our present study of the effects of support-configuration on the acceleration-sensitivity of the resonators, we chose an AT-cut plate of constant diameter-to-thickness ratio and subject to an in-plane acceleration of 15g. Therefore $\theta = 35^\circ 15'$, $R/b = 8.88$, and $G = 15\mu\text{g}$ and (1) may now be reduced to, for $n = 4$

$$\frac{\Delta f}{f_0} = F(\psi, \alpha_1, \alpha_2, \alpha_3, \alpha_4) \quad (2)$$

where α_i ($i = 1, 2, 3, 4$) control the configuration of supports. At a chosen reference support-configuration, the plate is supported by four metal ribbons which are 90° apart with two of them on the x_1 axis and the other two on the x_3 axis. Hence we have $\alpha_1 = 0^\circ$, $\alpha_2 = 90^\circ$, $\alpha_3 = 180^\circ$, and $\alpha_4 = 270^\circ$ for the reference configuration.

For a systematic study, seven series of configuration variations are considered. They are described as follows.

Series I: The support-configuration is varied from the reference configuration by displacing the support originally on the x_1 axis in increment of 15° counter-clockwise about the x_2 axis while holding the remaining three supports fixed. Therefore described in terms of α_i , we have $\alpha_2 = 90^\circ$, $\alpha_3 = 180^\circ$, $\alpha_4 = 270^\circ$, $\alpha_1 = 0^\circ + m \times 15^\circ$, $m = 1, 2, 3, \dots$ (See Fig. 2).

Series II: Similar to the Series I, except that it is the support originally on the $-x_3$ axis being displaced. Therefore $\alpha_1 = 0^\circ$, $\alpha_2 = 90^\circ$, $\alpha_3 = 180^\circ$, $\alpha_4 = 270^\circ + m \times 15^\circ$, $m = 1, 2, 3, \dots$ (See Fig. 3).

Series III: $\alpha_1 = 0^\circ$, $\alpha_3 = 180^\circ$, $\alpha_2 = 5^\circ + m \times 15^\circ$, $\alpha_4 = -\alpha_2$, $m = 0, 1, 2, \dots$ (See Fig. 4).

Series IV: $\alpha_2 = 90^\circ$, $\alpha_4 = 270^\circ$, $\alpha_1 = 265^\circ - m \times 15^\circ$, $\alpha_3 = 275^\circ + m \times 15^\circ$, $m = 1, 2, \dots$ (See Fig. 5).

Series V: $\alpha_1 = 0^\circ$, $\alpha_4 = 270^\circ$, $\alpha_2 = 127^\circ - m \times 30^\circ$, $\alpha_3 = 143^\circ + m \times 30^\circ$, $m = 0, 1, 2, 3, \dots$ (See Fig. 6).

Series VI: $\alpha_3 = 180^\circ$, $\alpha_4 = 270^\circ$, $\alpha_1 = 35^\circ - m \times 30^\circ$, $\alpha_2 = 55^\circ + m \times 30^\circ$, $m = 0, 1, 2, 3, \dots$ (See Fig. 7).

Series IIV: $\alpha_3 = 180^\circ$, $\alpha_4 = 270^\circ$, $\alpha_1 = 15^\circ + m \times 30^\circ$, $\alpha_2 = \alpha_1 + 90^\circ$, $m = 0, 1, 2, 3, \dots$ (See Fig. 8).

The changes in the resonance frequencies are computed as a function of the acceleration direction ϕ for each variation of the support-configuration of the seven Series described above.

Results and Conclusions

Calculations for frequency changes are made for a circular AT-cut quartz plate with one of the digonal axes of the crystal as the x_1 -axis of the coordinate system. The diameter-to-thickness ratio of the plate is chosen to be $R/b = 8.88$. The plate is subject to an acceleration of 15g and supported by four nickel ribbons of length $\ell = 6.35\text{mm}$ and rectangular cross-section ($h_1 = 0.076\text{mm}$, and $h_2 = 1.270\text{mm}$). The Young's modulus for Nickel is $E = 4.82 \times 10^9 \text{ dyne/cm}^2$.

Frequency changes are calculated as a function of the acceleration direction for each configuration variation. Results are presented as $\Delta f/\Delta f_{\text{max}}$ vs. ϕ in Figs. 2-8 corresponding to configuration variations of Series I-VII, respectively. Δf_{max} , chosen as the nondimensional factor for frequency changes, is the maximum frequency change in a given Series of configuration variations.

It is of interest to note that in each $\Delta f/\Delta f_{\text{max}}$ vs. ϕ curve there exists points (directions of acceleration) of relative maximum or minimum of frequency changes and points of zero frequency changes. In the neighborhood of the point of maximum frequency changes, Δf is usually less sensitive to the changes in the acceleration direction (i.e. $d(\Delta f/\Delta f_{\text{max}})/d\phi$ is small), while in the vicinity of the point of zero frequency changes the value of Δf is quite sensitive to the changes of acceleration directions. The point of maximum frequency changes and the point of zero frequency changes are usually spaced approximately 90° apart from each other.

The effect of support-configuration variations on the frequency changes can be seen by inspecting a sequence of $\Delta f/\Delta f_{\text{max}}$ vs. ϕ curves in a given Series. For instance, it may be seen in Fig. 2 for the configuration variations of Series I that the maximum frequency changes occur at $\alpha_1 = 60^\circ$, and for acceleration directions $\phi = 102^\circ$, 282° , and the directions for zero frequency changes occur at $\phi = 21^\circ$, 201° .

In Series I, II, VII (Fig. 2, 3 and 8, respectively) the locations of the points of maximum and zero frequency changes are effected by the changes of support-configuration. In Series III and IV (Figs. 4 and 5, respectively), directions of acceleration for maximum and zero frequency changes are independent of configuration changes, however the magnitude of the frequency changes is effected.

Although only seven Series of simple support-configuration variations are considered, however through the present parametric study one is able to see how the frequency changes due to the in-plane accelerations are influenced by the configuration variations.

References

1. P. C. Y. Lee, Kuang-Ming Wu, and Y. S. Wang, Research report No. 77-SM-6, Department of Civil Engineering, Princeton University, Princeton, N.J. (1977). Also Proc. 30th Annu. Freq. Control Symp. 1-8, (1976).
2. P. C. Y. Lee, Y. S. Wang, and X. Markenscoff, J. Acoust. Soc. Am. 57, 95-105 (1975). Also Proc. 27th Annu. Freq. Control Symp. 1-6, (1973).
3. A. W. Warner, Interim Reps. 10-11, Bell Telephone Lab. (1959). And W. L. Smith, Interim Reps. 12-13 and Final Rep., Bell Telephone Lab. (1960).

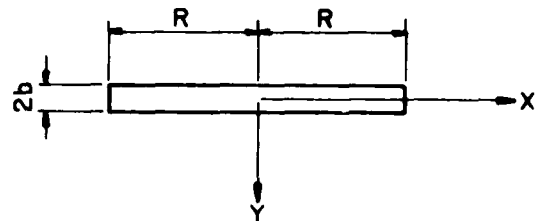
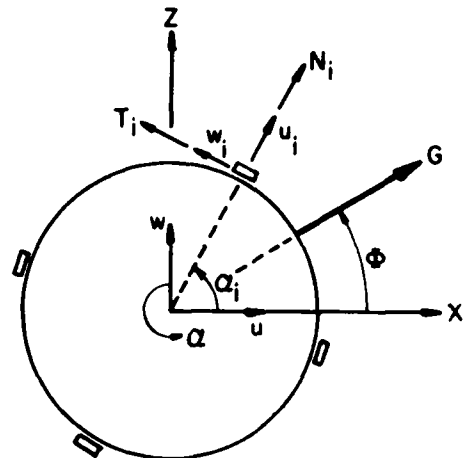


Figure 1 - A circular plate under body force G.

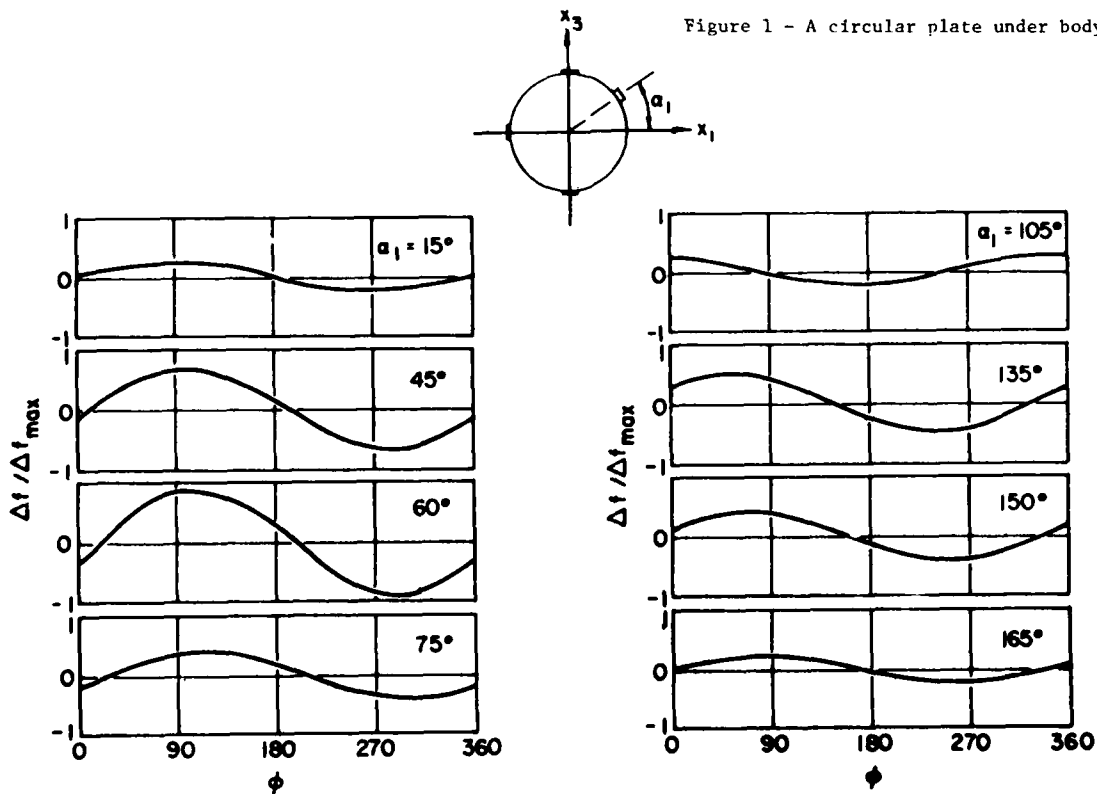


Figure 2 - Frequency changes $\Delta f / \Delta f_{\max}$ as a function of acceleration direction ϕ for the support configuration variation of Series I.

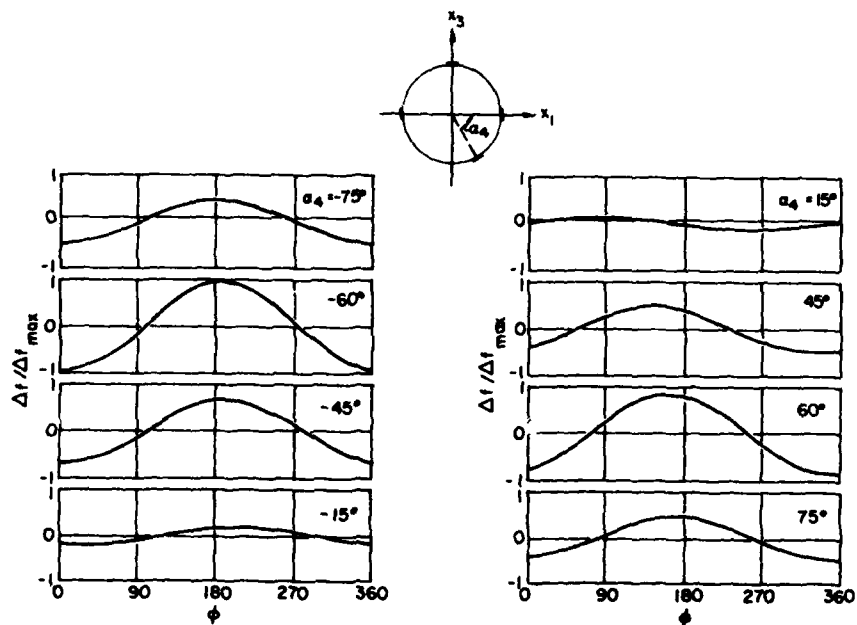


Figure 3 - Frequency changes $\Delta f/\Delta f_{\max}$ as a function of acceleration direction ϕ for the support-configuration variations of Series II.

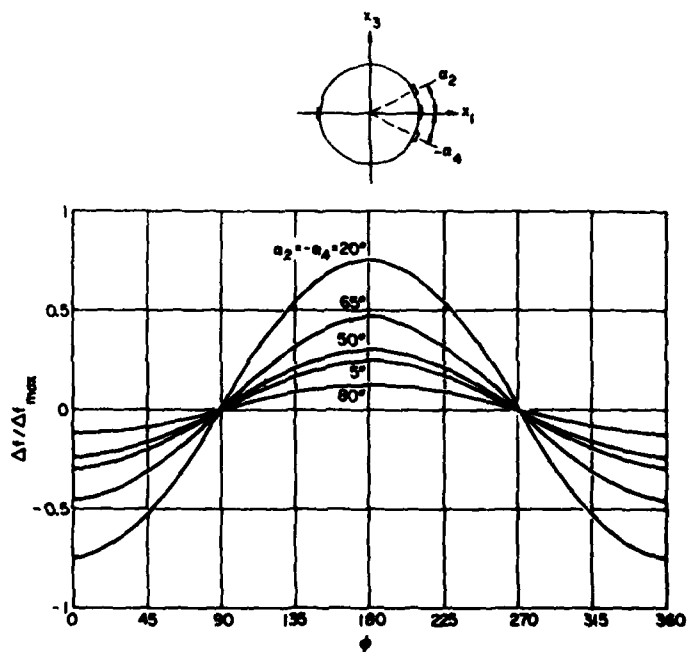


Figure 4 - Frequency changes $\Delta f/\Delta f_{\max}$ as a function of acceleration direction ϕ for the support configuration variation of Series III.

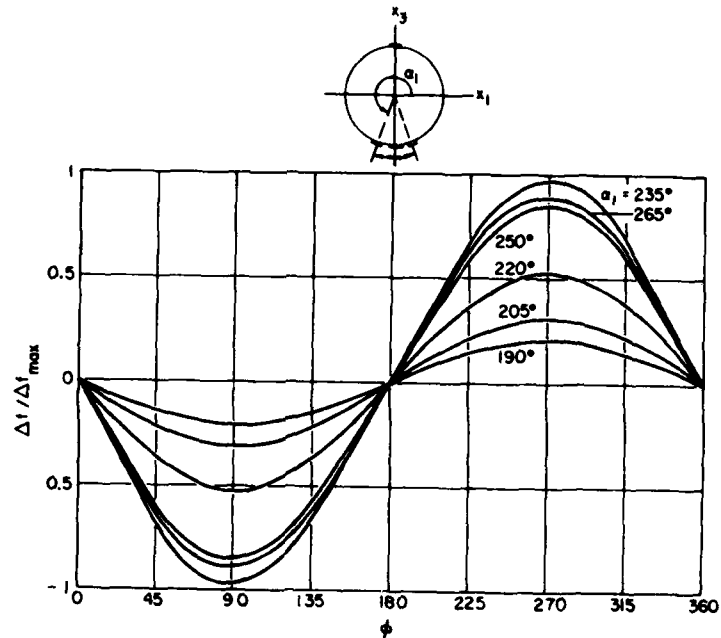


Figure 5 - Frequency changes $\Delta f / \Delta f_{\max}$ as a function of acceleration direction ϕ for the support configuration variation of Series IV.

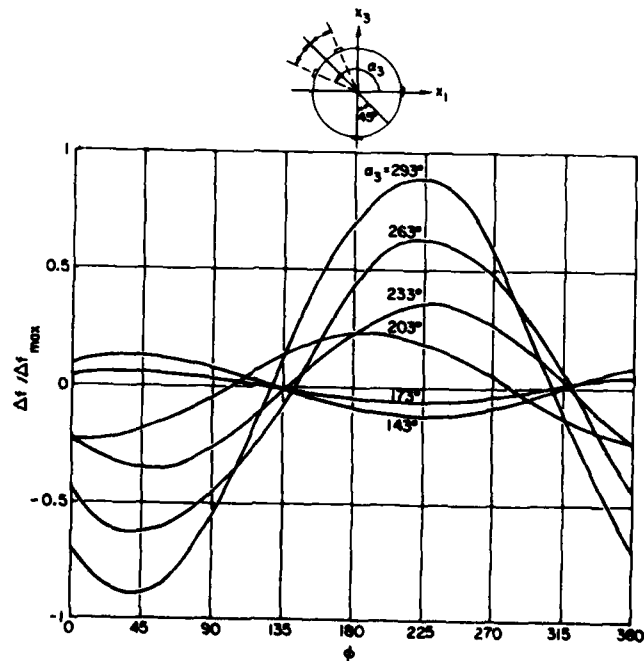


Figure 6 - Frequency changes $\Delta f / \Delta f_{\max}$ as a function of acceleration direction ϕ for the support configuration variation of Series V.

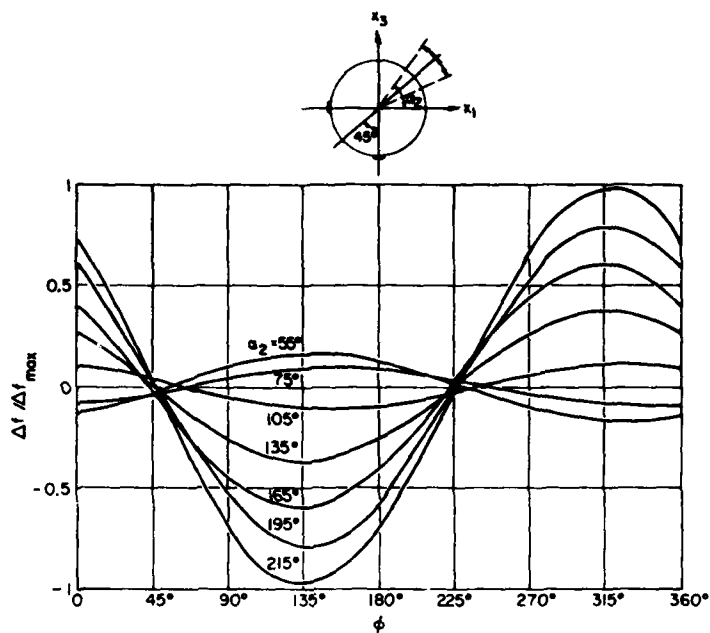


Figure 7 - Frequency changes $\Delta f/\Delta f_{\max}$ as a function of acceleration direction ϕ for the support configuration variations of Series VI.

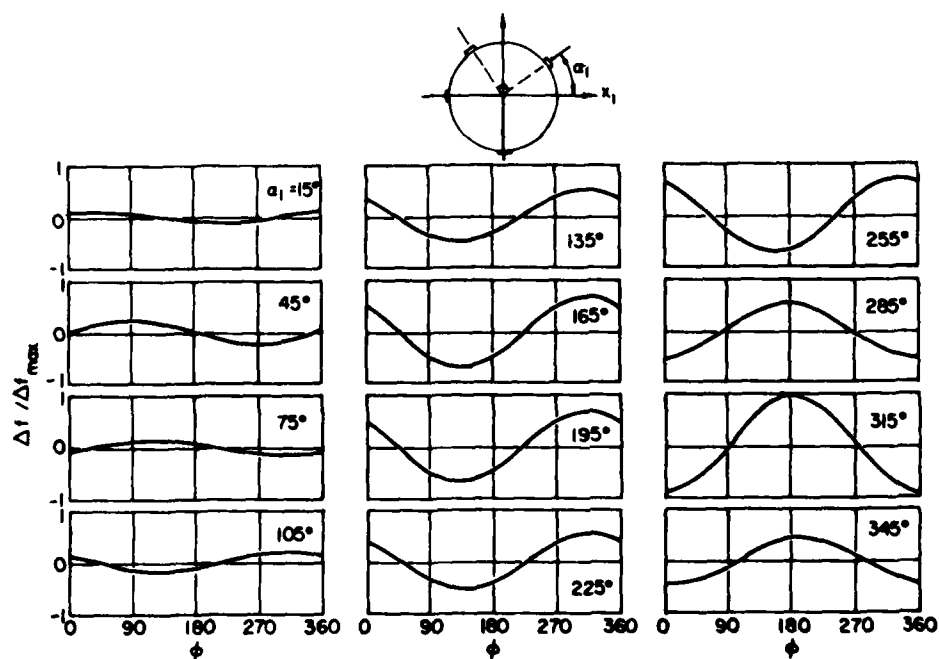


Figure 8 - Frequency changes $\Delta f/\Delta f_{\max}$ as a function of acceleration direction ϕ for the support configuration variations of Series VII.

X1 AND X3 FLEXURE, FACE-SHEAR, EXTENSION, THICKNESS-
SHEAR, AND THICKNESS-TWIST MODES IN
RECTANGULAR ROTATED Y-CUT QUARTZ PLATES

Thrygve R. Meeker
Bell Telephone Laboratories, Incorporated
Allentown, Pennsylvania

Summary

Simple transformations of variables that convert the characteristic matrix of the set of fully-coupled two dimensional differential equations for a rotated Y-cut quartz plate to a symmetric matrix are presented. The advantage of using this transformation is that efficient algorithms for obtaining the eigenvalues and eigenfunctions of such a system of equations are well-known. Scientific subroutine packages for large computing machines usually implement one or more of these algorithms. Material constants presented previously and the proposed transformation are used to calculate roots for the fully-coupled dispersion relations.

These dispersion relations allow C56 to have a finite value and allow both X1 and X3 variation at the same time. In these dispersion relations flexure, face-shear, extension, thickness-shear, and thickness-twist in both X1 and X3 directions are all coupled together. In order to visualize the meaning of these dispersion relations, simple approximate boundary conditions are used to determine the modes of a rectangular rotated Y-cut quartz plate. A large number of plate modes result from the different number of stress half-wavelengths along each direction for each of the five mode types. These modes may become unwanted resonances in rotated Y-cut quartz resonators.

Introduction

Unwanted resonances often limit the usefulness of crystal resonators. An oscillator circuit may lock on one of the unwanted resonances at some temperature or drive level, producing erratic shifts in the output frequency. A weaker unwanted mode may extract energy from the wanted mode and the resulting drop in resonator Q may cause a loss of oscillation. A crystal filter may transmit too much signal at the unwanted mode frequency in the filter stop band or the energy extracted from the desired mode may cause undesired changes in the dependence of pass band transmission on frequency.

The importance of understanding the complete frequency spectra of crystal resonators has resulted in many theoretical and experimental studies¹. One of the more useful theoretical developments has been the reduction of the three dimensional differential equations of motion, constitutive relations, and boundary conditions for an elastic plate to a truncated set of two dimensional equations^{2,3}. Similar two dimensional equations have been derived for the piezoelectric plate⁴. Both of these theoretical developments include the effects of crystal anisotropy. Flexure, face-shear, extension, thickness-twist, and thickness-shear modes are all coupled together. Thickness-stretch is not included in these equations. The truncated two dimensional equations have been used in a number of studies of wanted and unwanted modes in crystal plates. A review of all of these studies is beyond the scope of this paper. Most of these studies report various simplifications of the two dimensional

equations to reduce the computational complexity of a complete solution. These simplifications usually amount to setting various small constants or terms in the equations to zero to decouple two or three of the modes and then solving the resulting equations.

Three of the studies mentioned above are particularly relevant to the purpose of the present paper. Mindlin's two dimensional equations for the anisotropic crystal plate with rotatory inertia and transverse shear deformation were reduced to the specific case of rotated Y-cut quartz by Mindlin and Spencer⁵. These general equations were then simplified by removing thickness-shear, dropping inertia from some of the equations, letting the shear strain component ϵ_{24} approach zero while keeping a finite zeroth order stress component $T_{23}(0)$, and setting several small material constants (γ_{13} , γ_{33} , and C56) equal to zero in some equations. The resulting equations had flexure and thickness-twist coupled together, but these two coupled modes were uncoupled from thickness-shear, extension, and face-shear. These flexure-twist equations were solved for a rectangular elastic plate with stress free edges. Agreement with measured modes in quartz resonators was quite good⁶. Lee⁶ has used similar general equations to study coupled extensional, flexural, and width shear modes in 5° X-cut quartz plates. Good agreement with experiment was found over a wide frequency range. Lee and Spencer⁷ studied the shear-flexure-twist vibrations of a partly electroded rotated Y-cut quartz plate. Agreement with experiment was also very good.

These examples demonstrate the general ability of Mindlin's 1961 equations to predict the frequencies of various modes in quartz crystal plates. These and other published reports also suggest that substantial changes in the Mindlin equations can be made without losing good agreement between theory and experiment. Tiersten has reduced the general set of piezoelectric two dimensional equations to the case of rotated Y-cut quartz⁸.

Details in procedures for the numerical evaluation of these equations have not usually been reported. The present paper discusses the numerical evaluation of the dispersion relation of Tiersten's equations for rotated Y-cut quartz. The electrical equations and piezoelectric terms are dropped. No additional simplifications are used. Although the resulting equations have the same form as the Mindlin-Spencer equations, the elastic constants are slightly different because electrode stiffening is included. Dispersion relations for X1 dependence or X3 dependence alone were reported by Mindlin and Spencer⁵. In the present paper, X1 and X3 dependence are allowed at the same time. Some implications of this two dimensionality will be discussed. Mindlin and Spencer suggest that the full set of five equations do not have any solutions in closed form.

In the present paper approximate boundary conditions which remove this limitation are presented. The resulting frequency equation retains all of the

appropriate material constants and all of the five differential equations. Extension, flexure, face-shear, thickness-shear, and thickness-twist, with X_1 and X_3 dependence at the same time, are all coupled together in these results. Electrode elasticity and mass are included. A comparison of experimental results with the predicted frequencies shows the same kind of agreement as discussed above for other situations.

Differential Equations of Motion

The developments reported in this paper begin with the general two dimensional equations for rotated Y-cut quartz as given by Tiersten⁸. The electrical equations and the electrical terms in the elastic equations are dropped. Tiersten has shown how to include the effects of electrode elasticity in the two dimensional constants⁹. Techniques for evaluating the zero order and first order two dimensional elastic constants for rotated Y-cut quartz plates have been described before¹⁰. Electrode elasticity effects are included by using Tiersten's Equation 14.32¹¹.

$$C_{ijkl} = C_{ijkl}^{(u)} + R [(\lambda^0 \delta_{ib} \delta_{jb} \delta_{kd} \delta_{ld}) + \mu (\delta_{ic} \delta_{jd} \delta_{kc} \delta_{ld} + \delta_{id} \delta_{jc} \delta_{kc} \delta_{ld})] \quad (1)$$

and

$$\gamma_{abcd} = \gamma_{abcd}^{(u)} + 3R [\lambda^0 \delta_{ab} \delta_{cd} + \mu (\delta_{ac} \delta_{bd} + \delta_{ad} \delta_{bc})], \quad (2)$$

where $\lambda^0 = 2\mu \lambda / (\lambda + 2\mu)$ and $R = 2 \rho_1 h_1 / \rho h$.

$C_{ijkl}^{(u)}$ and $\gamma_{abcd}^{(u)}$ are the zeroth order elastic constants for the unelectroded plate. λ and μ are the Lamé constants for the isotropic electrode material. ρ_1 and h_1 are the density and thickness of each of the identical electrodes; ρ and h are the density and thickness of the quartz; and δ_{ij} is the Kronecker delta.

The differential equations of motion then take the following form:

Equations 3 through 7 are five coupled second order differential equations in the three zeroth order and two first order displacement components of the Mindlin theory². The form of these five components is shown in Figure 1¹². These components are the coefficients of the expansion of the particular displacement in a power series in X_2^2 , which is the thickness direction of the plate. Although the displacement components do not depend on X_2^2 , they may depend on X_1 and X_3^2 as necessary to satisfy Equations 3 through 7 and the appropriate boundary conditions.

Arbitrary constants K_4 and K_6 are adjusted to make the thickness-shear and thickness-twist frequencies of the infinitely extended plate equal to the exact frequencies of the three dimensional plate. These constants are K_1 and K_3 in Mindlin and Spencer's notation. K_6 (or K_1) matches the thickness-twist frequency and K_4 (or K_3) matches the thickness-shear frequency. An approximate value of K_6^2 was given by Mindlin as $\pi^2/12$ but it has become customary to solve for the infinite plate frequency exactly and set K_6 so that the two dimensional frequency and the three dimensional frequency agree exactly for the infinite plate. K_4 is determined from K_6 and the elastic constants of the plate, as

$$K_4^2 = K_6^2 \frac{[C_{22}^E + C_{44}^E - ((C_{22}^E - C_{44}^E)^2 + 4C_{24}^E)^{1/2}]}{2C_{44}} \quad (8)$$

The C_{ij}^E in the numerator of Equation 8 are the rotated three dimensional elastic constants; the C_{44} in the denominator is the two dimensional elastic constant as stiffened by the electrode.

Equations 3 through 7 have the same form as Mindlin and Spencer's Equations 4 and 5⁵. The elastic constants (C_{ij} and γ_{ij}) in the Tiersten equations used here are slightly different because electrode elasticity effects are included. Electrode mass effects are included in both Tiersten and Mindlin-Spencer by multiplying the density by $1 + R$ or $1 + 3R$ as indicated.

$$C_{11}^{(0)} u_{1,1}^{(0)} + C_{13}^{(0)} u_{3,1}^{(0)} + K_4 C_{14} (u_{2,1}^{(0)} + u_{3,1}^{(1)}) + C_{55}^{(0)} (u_{3,1}^{(0)} + u_{1,3}^{(0)}) + K_6 C_{56} (u_{2,1}^{(0)} + u_{1,3}^{(1)}) = \rho (1+R) \ddot{u}_1^{(0)} \quad (3)$$

$$C_{13}^{(0)} u_{1,1}^{(0)} + C_{33}^{(0)} u_{3,1}^{(0)} + K_4 C_{34} (u_{2,1}^{(0)} + u_{3,1}^{(1)}) + C_{55}^{(0)} (u_{3,1}^{(0)} + u_{1,3}^{(0)}) + K_6 C_{56} (u_{2,1}^{(0)} + u_{1,3}^{(1)}) = \rho (1+R) \ddot{u}_3^{(0)} \quad (4)$$

$$K_4 C_{14} u_{1,1}^{(0)} + K_4 C_{34} u_{3,1}^{(0)} + K_4^2 C_{44} (u_{2,1}^{(0)} + u_{3,1}^{(1)}) + K_6 C_{56} (u_{3,1}^{(0)} + u_{1,3}^{(0)}) + K_6^2 C_{66} (u_{2,1}^{(0)} + u_{1,3}^{(1)}) = \rho (1+R) \ddot{u}_2^{(0)} \quad (5)$$

$$\gamma_{11}^{(1)} u_{1,1}^{(1)} + \gamma_{13}^{(1)} u_{3,1}^{(1)} + \gamma_{55}^{(1)} (u_{3,1}^{(1)} + u_{1,3}^{(1)}) - 3b^{-2} [K_6 C_{56} (u_{3,1}^{(0)} + u_{1,3}^{(0)}) + K_6^2 C_{66} (u_{2,1}^{(0)} + u_{1,3}^{(1)})] = \rho (1+3R) \ddot{u}_1^{(1)} \quad (6)$$

$$\gamma_{55}^{(1)} (u_{3,1}^{(1)} + u_{1,3}^{(1)}) + \gamma_{13}^{(1)} u_{1,1}^{(0)} + \gamma_{13}^{(1)} u_{3,1}^{(1)} - 3b^{-2} [K_4 (C_{14}^{(0)} u_{1,1}^{(0)} + C_{34}^{(0)} u_{3,1}^{(0)}) + K_4^2 C_{44} (u_{2,1}^{(0)} + u_{3,1}^{(1)})] = \rho (1+3R) \ddot{u}_3^{(1)} \quad (7)$$

Solutions to the Differential Equations of Motion

A popular method¹³ of solving the equations of motion is to assume a standing wave solution of the form

$$Z = Z_0 \begin{Bmatrix} \sin k_1 x_1 \\ \cos k_1 x_1 \end{Bmatrix} \begin{Bmatrix} \sin k_3 x_3 \\ \cos k_3 x_3 \end{Bmatrix} \sin \omega t, \quad (9)$$

where $Z = u_1^{(0)}, u_3^{(0)}, u_2^{(0)}, u_1^{(1)},$ and $u_3^{(1)}$ and $Z_0 = u_{10}, u_{30}, u_{20}, s_{10},$ and s_{30} . Four solutions with different combinations of the trigonometric functions are possible. Only solutions with real k_1 and k_3 will be discussed in this paper. The assumed wave functions are solutions of the differential equations of motion for particular values of k_1, k_3 and ω . The relationship among k_1, k_3 and ω so that the wave functions are solutions to the differential equations of motion is called the dispersion relation. One of the four possible wave functions³ is

$$u_1^{(0)} = u_{10} \cos k_1 x_1 \cos k_3 x_3 \sin \omega t = u_{10} C_1 C_3 S_1$$

$$u_3^{(0)} = u_{30} \sin k_1 x_1 \sin k_3 x_3 \sin \omega t = u_{30} S_1 S_3 S_1$$

$$u_2^{(0)} = u_{20} \sin k_1 x_1 \sin k_3 x_3 \sin \omega t = u_{20} S_1 S_3 S_1 \quad (10)$$

$$u_1^{(1)} = s_{10} \cos k_1 x_1 \sin k_3 x_3 \sin \omega t = s_{10} C_1 S_3 S_1$$

$$u_3^{(1)} = s_{30} \sin k_1 x_1 \cos k_3 x_3 \sin \omega t = s_{30} S_1 C_3 S_1$$

Equation 10 is a solution to the five differential equations of motion (Equations 3 through 7) if

$$\{ [D] - \rho \omega^2 [B] \} [\phi] = [0], \quad (11)$$

where

$$[D] = \begin{bmatrix} (c_{11}k_1^2 + c_{55}k_3^2) (-c_{13}k_1k_3 - c_{55}k_1k_3) (-K_4c_{14}k_1k_3 - K_6c_{56}k_1k_3) (-K_6c_{56}k_3 \frac{\sqrt{3}}{b}) (-K_4c_{14}k_1 \frac{\sqrt{3}}{b}) \\ D_{21} \quad (c_{33}k_3^2 + c_{55}k_1^2) \quad (K_4c_{34}k_3^2 + K_6c_{56}k_1^2) \quad (K_6c_{56}k_1 \frac{\sqrt{3}}{b}) \quad (K_4c_{34}k_3 \frac{\sqrt{3}}{b}) \\ D_{31} \quad D_{32} \quad (K_4^2c_{44}k_3^2 + K_6^2c_{66}k_1^2) \quad (K_6^2c_{66}k_1 \frac{\sqrt{3}}{b}) \quad (K_4^2c_{44}k_3 \frac{\sqrt{3}}{b}) \\ D_{41} \quad D_{42} \quad D_{43} \quad (\gamma_{11}k_1^2 + \gamma_{55}k_3^2 + \frac{3}{b^2}K_6^2c_{66}) \quad (\gamma_{13}k_1k_3 + \gamma_{55}k_1k_3) \\ D_{51} \quad D_{52} \quad D_{53} \quad D_{54} \quad (\gamma_{55}k_1^2 + \gamma_{33}k_3^2 + \frac{3}{b^2}K_4^2c_{44}) \end{bmatrix} \quad (12)$$

$$\theta_{ij} = \theta_{1j} [1 + \theta_{2j} 3R + \theta_{3j} 3R + \theta_{4j} R + \theta_{5j} R] \quad (13)$$

and

$$[\phi] = \begin{bmatrix} u_1^0 \\ u_3^0 \\ u_2^0 \\ \psi_{10} \\ \psi_{30} \end{bmatrix} \quad (14)$$

$$\text{with } \psi_{10} = \frac{b}{\sqrt{3}} s_{10} \text{ and } \psi_{30} = \frac{b}{\sqrt{3}} s_{30}. \quad (15)$$

Using ψ_{10} and ψ_{30} as wave function component amplitudes, rather than s_{10} and s_{30} , makes D symmetric. Abbreviations used in this paper for the trigonometric functions (S_1 , S_3 , C_1 , C_3 , and S_t) are defined in Equation 10. For a given value of ω and k_1 (or k_3) there are five values of k_3 (or k_1) for which the five homogeneous equations (Equation 11) have a solution, with $\det(D - \rho\omega^2 B) = 0$. Each of the five solutions is associated with a different set of ratios of the five components of the displacement. In the usual solution sequence¹³ these five displacement solutions are linearly combined in a general wave function. This wavefunction and the boundary conditions are substituted into the constitutive equations. For stress free or strain free boundary conditions, the result is a set of homogeneous equations, often called the frequency equations. The frequencies for which a solution exists are the resonant frequencies of the system. For x_1 and x_3 dependence the above process usually needs one root search (on k_1 or k_3) to solve the dispersion relation and a second (on ω) to solve the frequency equation.

The main purpose of the present paper is to discuss an approximation that makes both root searches unnecessary. The resulting approximate solution has substantial advantages in computing time and cost, while providing considerable insight into the properties of many of the modes of the plate. In this approximation the dispersion relation for the complete problem is solved exactly by using ω as a dependent variable instead of an independent variable. A wave-number substitution converts the exact dispersion relation into an approximate frequency equation. The resulting frequency equation is solved exactly. For this paper only the frequency equation is solved. Equation 11 is already in the proper form for the present solution.

In Equation 11 $[D]$ is a function of elastic constants and wavenumbers k_1 and k_3 . If an exact solution for the dispersion relation with real wavenumbers is desired, real values of k_1 and k_3 are substituted into $[D]$. To convert Equation 11 into an approximate frequency equation, k_1 and k_3 are chosen so that the plate width and length are each an integral number of half wavelengths. Then,

$$k_1 = \frac{2\pi}{\lambda_1} = \frac{2\pi}{2W/n_1} = n_1 \frac{\pi}{W}, \quad (16)$$

$$\text{and } k_3 = \frac{2\pi}{\lambda_3} = \frac{2\pi}{2L/n_3} = n_3 \frac{\pi}{L}, \quad \text{where} \quad (17)$$

n_1 is the number of half wavelengths in the x_1 direction and n_3 is the number of half wavelengths in the

x_3 direction. This approximation will be discussed in more detail in a later section of this paper.

After values of k_1 and k_3 have been assigned, all the elements of D can be calculated. Equation 11 becomes an eigenvalue problem, in which the eigenvalues are simple functions of frequency. The symmetry of D , generated by the two simple transformations in Equation 15, guarantees real eigenvalues and frequencies. In addition, many computing machines have stored programs which solve for eigenvalues and eigenvectors of symmetric real coefficient systems¹⁴.

$$\text{Then } [D] [\phi^V] = [\lambda] [\phi^V], \quad (18)$$

where $[\lambda]$ is a diagonal matrix with real elements and $[\phi^V]$ has the same dimension as $[D]$. The diagonal elements of $[\lambda]$ will be called λ_j . Each column of $[\phi^V]$ is an eigenvector, associated with one of the diagonal elements of $[\lambda]$. Each column of $[\phi^V]$ has the same elements as $[\phi]$, but the element ratios are different for each eigenvector. The particular algorithm used here outputs the eigenvalues in order of decreasing value. Each eigenvector is associated with an eigenvalue and is normalized to unit amplitude.

$$\text{Since } \{ [D] - \rho\omega^2 [B] \} [\phi^V] = [0] \quad (19)$$

$$\text{by Equation 11, } \rho\omega^2 [B] = [\lambda]. \quad (20)$$

$$\text{Then } \omega_j^2 = \frac{\lambda_j}{(1+3R)\rho} \quad \text{for } j = 1, 2 \quad (21)$$

$$\text{and } \omega_j^2 = \frac{\lambda_j}{(1+R)\rho} \quad \text{for } j = 3, 4, 5, \quad (22)$$

as long as thickness-shear and thickness-twist have the highest frequencies. Equations 21 and 22 must be modified if this usual condition is not met. The eigenvector component values show the relative proportions of the five wavefunction amplitude components associated with each eigenvalue. In general, the eigenvector motions have all five amplitude components. Under some conditions one component dominates.

The Wavenumber Approximation

Setting a dimension equal to an integral number of wavelengths to derive a frequency equation has been used for a long time to obtain one dimensional approximations to the frequency of vibrating systems. For example, McSkimin¹⁵ discusses such an approximation. In the present approximation there are five wavelengths in each of two directions instead of one wavelength in one direction. But the principle is the same. The modes arising from the present approximation also differ slightly from the Lamé type modes derived by Mindlin¹⁶ because no restrictions on crystal orientation or dimensions are needed.

Substituting the displacement solution of Equation 10 into Tiersten's constitutive equations¹⁷ gives the following:

$$T_{11}^{(0)} = 2b \left[-c_{11} k_1 u_{10} + c_{13} k_3 u_{30} + K_4 c_{14} k_3 u_{20} + K_4 c_{14} s_{30} \right] S_1 C_3 S_1 \quad (23)$$

$$T_{33}^{(0)} = 2b \left[-c_{13} k_1 u_{10} + c_{33} k_3 u_{30} + K_4 c_{34} k_3 u_{20} + K_4 c_{34} s_{30} \right] S_1 C_3 S_1 \quad (24)$$

$$T_{13}^{(0)} = 2b \left[-c_{55} k_3 u_{10} + c_{55} k_1 u_{30} + K_6 c_{56} k_1 u_{20} + K_6 c_{56} s_{10} \right] C_1 S_3 S_1 \quad (25)$$

$$T_{12}^{(0)} = 2b \left[-K_6 c_{56} k_3 u_{10} + K_6 c_{56} k_1 u_{30} + K_6^2 k_1 u_{20} + K_6^2 c_{66} s_{10} \right] C_1 S_3 S_1 \quad (26)$$

$$T_{23}^{(0)} = 2b \left[-K_4 c_{14} k_1 u_{10} + K_4 c_{34} k_3 u_{30} + K_4^2 c_{44} k_3 u_{20} + K_4^2 c_{44} s_{30} \right] S_1 C_3 S_1 \quad (27)$$

$$T_{11}^{(1)} = \frac{2}{3} b^3 \left[-\gamma_{11} k_1 s_{10} - \gamma_{13} k_3 s_{30} \right] S_1 S_3 S_1 \quad (28)$$

$$T_{33}^{(1)} = \frac{2}{3} b^3 \left[-\gamma_{13} k_1 s_{10} - \gamma_{33} k_3 s_{30} \right] S_1 S_3 S_1 \quad (29)$$

$$T_{13}^{(1)} = \frac{2}{3} b^3 \left[+\gamma_{55} k_3 s_{10} + \gamma_{55} k_1 s_{30} \right] C_1 C_3 S_1 \quad (30)$$

where $S_1 = \sin(k_1 x_1)$, $C_1 = \cos(k_1 x_1)$, $S_3 = \sin(k_3 x_3)$, $C_3 = \cos(k_3 x_3)$, and $S_t = \sin \omega t$. If the coordinate system of the plate is selected so that one corner is at $x_1 = 0$ and $x_3 = 0$, the definitions of Equation 16 and Equation 17 make $S_1 = 0$ for $x_1 = W$ for all integer values of n_1 and make $S_3 = 0$ for $x_3 = 0$ and for $x_3 = L$ for all integer values of n_3 .

The approximations of Equations 16 and 17 then amount to setting

$$T_{11}^{(0)}, T_{33}^{(0)}, T_{23}^{(0)}, T_{31}^{(0)}, T_{13}^{(0)}, T_{11}^{(1)}, T_{33}^{(1)}, u_3^{(0)}, u_2^{(0)}, \text{ and } u_3^{(1)} \text{ all}$$

equal to zero on $x_1 = 0$ and $x_1 = W$ and to setting

$$T_{13}^{(0)}, T_{12}^{(0)}, T_{11}^{(0)}, T_{31}^{(0)}, T_{33}^{(1)}, T_{31}^{(1)}, u_3^{(0)}, u_2^{(0)} \text{ and } u_1^{(1)} \text{ all equal to zero on}$$

$x_3 = 0$ and $x_3 = L$. The uniqueness theorem¹⁸ for this problem only requires that one member from each of the following products be specified on the boundaries:

$$T_{11}^{(0)} u_1^{(0)}, T_{13}^{(0)} u_3^{(0)}, T_{31}^{(0)} u_1^{(0)}, T_{12}^{(0)} u_2^{(0)}, T_{32}^{(0)} u_2^{(0)}, T_{33}^{(0)} u_3^{(0)}, T_{11}^{(1)} u_1^{(1)}, T_{31}^{(1)} u_1^{(1)}, T_{33}^{(1)} u_3^{(1)},$$

and $T_{13}^{(1)} u_3^{(1)}$.

The approximation of using Equations 16 and 17 then amounts to using mixed boundary conditions, i.e., some $T_{ij}(n) = 0$ and some $u_j(n) = 0$. Since most of the $T_{ij}(n) = 0$ on the appropriate plate edge, the resulting calculated frequencies should be close to observed frequencies, except at frequencies for which strong coupling occurs at the plate boundaries. No boundary coupling is present in the approximation.

Results

Although all of the results to be discussed here are presented in terms of resonant frequencies, the plots are exact solutions for the dispersion relations for the appropriate values of k_1 and k_3 . The present approximation only affects how values of k_1 and k_3 are selected. If a more exact frequency equation were desired, according to the usual method of using linear combinations of solution vectors with the same frequency, it would become necessary to perform the two root searches mentioned earlier. It is not known whether it is more efficient to assign values of k_1 and ω in Equation 11 and then vary k_3 for a zero in the determinant of the coefficients of the homogeneous equation, repeating the process for the desired number of values of k_1 , or to assign desired values of k_1 and vary k_3 for each value until the frequency equals the desired frequency.

In this section a few examples of results obtained with the present approximation are discussed. Figure 2 shows calculated resonant frequencies versus the number (n_3) of half wavelengths along the x_3 or Z' direction of the AT quartz plate. The unelectroded plate has a thickness near .0002056 m., such that the thickness-twist frequency (F_c) for infinite lateral dimensions ($n_1 = 0$, $n_3 = 0$) is 8.0 MHz. Thickness-shear (TS_3), thickness-twist (TT_3), extensional (E_3), face-shear (FS_3) and flexural (F_3) modes are shown.

These results are the same as those presented by Mindlin and Spencer in their Figure 1⁵. In Figure 2

the logarithm of the resonant frequency is plotted to show some of the detail of the lower frequency contour modes on the same plot as the high frequency modes. For the case of $n_1 = 0$ (or $n_3 = 0$, not shown) the eigenvectors each have one component, as marked on the curves. Frequencies for up to 100 half wavelengths are shown. F_3 , FS_3 , and E_3 modes all have zero frequency for $n_3 = 0$. According to Equation 17, $n_3 = 0$ results from $L = \infty$. TI_3 and TS_3 modes have cutoff frequencies at $L = \infty$.

The fundamental frequency for each of the five mode types is the case for $n_1 = 1$ and $n_3 = 1$, and the frequencies for $n_1 > 1$, $n_3 > 1$ are the inharmonic frequencies of the modes.

Because the modes do not usually have single component eigenvectors, as in Figure 2, it is better to name the modes in order of decreasing frequency, as $F(3)$ (n_1, n_3). $F(5)$ is usually a flexure, $F(1)$ is usually thickness-shear, and $F(2)$ is usually thickness-twist. $F(3)$ and $F(4)$ are usually coupled face-shear and extension. These mode names are also shown in Figure 2. As pointed out by Mindlin and Spencer, similar results hold for the cylindrical modes with x_1 variation alone.

Figure 3 shows what happens to the calculated frequencies if there are 10 half wavelengths along the plate width ($n_1 = 10$). For this case, all five modes have cutoff frequencies (finite frequencies at infinite length). $F(4)$ ($10, n_3$) has a minimum frequency for n_3 between 10 and 20. Figure 4 shows calculated frequencies for some lower frequency modes on a linear frequency scale. In Figures 3 and 4 $F(5)$ is mostly flexure ($u_2^{(0)}$), but $F(3)$ and $F(4)$ are combinations of face-shear and extension ($u_1^{(0)}$ and $u_3^{(0)}$).

One use of the present results is to identify unwanted resonances in thickness-shear or thickness-twist resonators. Figure 5 shows the values of n_1 and n_3 for all modes that have calculated frequencies between 7.9 and 8.1 mHz. The thickness-twist mode for this plate is near 8.0 mHz. In this case there were 487 unwanted modes as indicated. n_1 was varied between 0 and 60 and n_3 was varied between 0 and 125. All combinations of n_1 and n_3 with $n_1 > 61$ and $n_3 > 125$ were outside of the indicated frequency range.

Eigenvectors for five points on this curve are shown on the figure. The multicomponent character of these eigenvectors can be seen from this figure. Points from 0,0 to 10,10 define inharmonic modes of thickness-shear and thickness-twist. The other points define high order extension, face-shear, and flexure modes. It is fortunate for the device application of thickness-shear and thickness-twist modes that most of the inharmonic shear and twist and the high order flexure, face-shear, and extensions are weakly excited. Some implications of the twist inharmonics and high order flexure and face-shear are discussed by Pearman and Rennick¹⁹.

Calculations such as those shown in Figure 5 have important consequences to the further development of the theory of modes in plates. The calculations in Figure 5 suggest that modes with $n_1 < 62$ and $n_3 < 127$ are the only ones with real wavenumber needed in an eigenvector expansion. This information could make the eigenvector expansion finite and therefore expressible as a closed form function. Further work may clarify how to use results such as Figure 5 in more exact approximations. In this case complex wavenumber values are also important.

Table I shows how some mode frequency separations calculated with the present approximation agree with experimental results published by Heising²⁰.

TABLE I
COMPARISON OF MODE SPACINGS
FOR AT QUARTZ PLATES

Mode	Reference	ΔF (KHz)
I Length $2B=.000208m$	Heising	125.0
Face	EQ. 6.25	
Shear $L=.02032 m$	This Paper	$F_{(4)}(1,65) - F_{(4)}(1,64) = 125.8$
$W=.02032 m$		
II Width $2b=.0010 m$	Heising	$F_{(38)} - F_{(37)} = 45.0$
Flexure	Fig. 6.14	$F_{(36)} - F_{(35)} = 50.0$
$L=.0320 m$	This Paper	$F_{(5)}(38,1) - F_{(5)}(37,1) = 48.5$
$W=.0320 m$		$F_{(5)}(36,1) - F_{(5)}(37,1) = 48.6$

The agreement between calculated and experimental length face-shear and width flexural modes is good enough to indicate the general usefulness of the present work.

Conclusions

Tiersten's two dimensional differential equations for AT-cut quartz with piezoelectric terms and equations dropped but including electrode mass and stiffness are solved exactly for the dispersion relation among frequency and wavenumbers in x_1 and x_3 directions for thickness-shear, thickness-twist, extension, face-shear, and flexure modes.

Simple linear transformations make the characteristic matrix of the dispersion relation symmetric. If the wavenumbers are made the independent variables, the resulting dispersion relation can be solved exactly without root searches by using standard eigenvalue-eigenvector techniques. A much used approximation in which plate length and width are set equal to an integral number of half wavelengths converts the exact dispersion relation to an approximate frequency equation. This frequency equation may also be solved directly by eigenvalue techniques without root searches. The present equations reduce to the Mindlin-Spencer cylindrical modes when there is no variation along x_3 or along x_1 . For arbitrary variation along both directions, inharmonic modes of all five mode types are found. Some branches have minima and all branches have cutoff frequencies under some conditions. Only real wavenumbers are studied in this paper. Calculated mode spacings for quartz AT plates are in reasonable agreement with experiment. The results of this study should indicate some directions of future theoretical developments.

References

1. R. A. Heising, "Quartz Crystals for Electrical Circuits", D. Van Nostrand Co., Inc., New York, 1946, Chapter VI, pps. 205-275 (R. A. Sykes).
2. R. D. Mindlin, "An Introduction to the Mathematical Theory of the Vibrations of Elastic Plates", U.S. Army Signal Corps Eng. Lab., Fort Monmouth, N.J. (1955), Signal Corps Contract DA-36-039 SC-56772.

3. R. D. Mindlin, Quart. Appl. Math 19, 51-61 (1961).
4. H. F. Tiersten, "Linear Piezoelectric Plate Vibrations", Plenum Press, New York, 1969, Chapter 13, pps. 141-168.
5. R. D. Mindlin and W. J. Spencer, J. Acoust. Soc. Amer. 42, 1268-1277 (1967).
6. P. C. Y. Lee, J. Appl. Phys. 42, 4139-4144 (1971).
7. P. C. Y. Lee and W. J. Spencer, J. Acoust. Soc. Amer. 45, 637-645 (1969).
8. H. F. Tiersten, *ibid.*, Chapter 13, Section 8, pps. 160-162.
9. H. F. Tiersten, *ibid.*, Chapter 14, 169-178.
10. T. R. Meeker, "Plate Constants and Dispersion Relations for Width-Length Effects in Rotated Y-Cut Quartz Plates", Proc. of 29th Annual Symposium of Frequency Control, U.S. Army Electronics Command, Fort Monmouth, N.J., pps. 54-64 (1975).
11. H. F. Tiersten, *ibid.*, Chapter 14, page 175.
12. A. A. Comparini and J. J. Hannon, IEEE Trans. on Sonics and Ultrasonics, Vol. SU-21, No. 2, pps. 130-135, April, 1974.
13. H. F. Tiersten, *ibid.*, Chapter 8, pps. 65-80.
14. F. S. Acton, "Numerical Methods that Work", Harper and Row, New York (1970), Chapter 8, pps. 204-220 and Chapter 13, pps. 316-360.
15. R. A. Heising, *ibid.*, p. 271 (H. J. McSkimin).
16. R. D. Mindlin, J. Appl. Phys. 27, pps. 1462-1466 (1956).
17. H. F. Tiersten, *ibid.*, pages 160-161.
18. H. F. Tiersten, *ibid.*, page 152.
19. G. T. Pearman and R. C. Rennick, to be presented at 31st Annual Frequency Control Symposium, 1977.
20. R. A. Heising, *ibid.*, pps. 218-237 (R. A. Sykes).

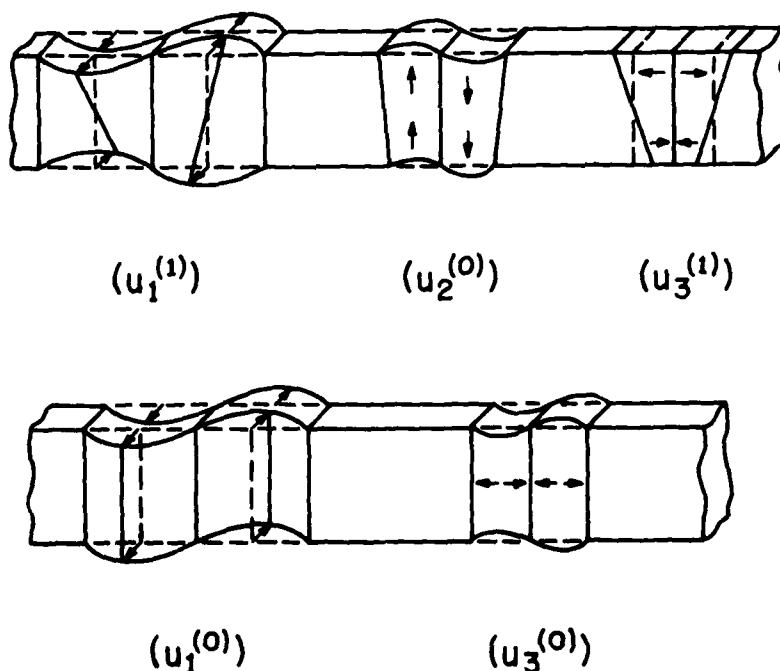


Figure 1 - Elementary Displacement Components

DEPENDENCE OF RESONANT FREQUENCY ON WAVE NUMBER

AT QUARTZ, $W = .0106\text{m}$, $L = .02032\text{m}$, $F_0 = 8.0\text{MHz}$, $T = 25^\circ\text{C}$, $n_1 = 0$

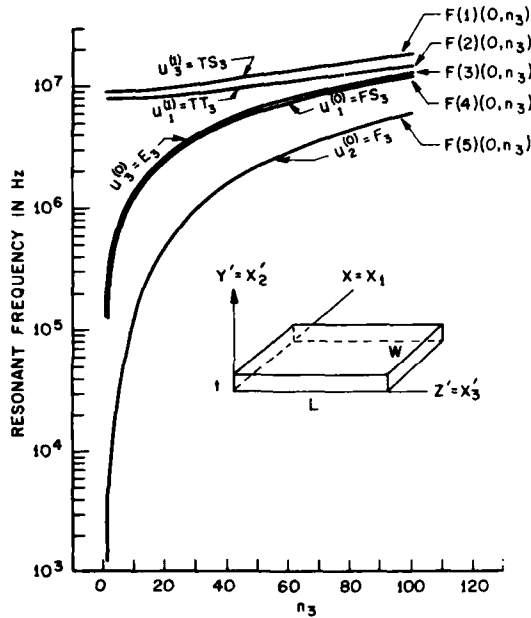


Figure 2

Dependence of Resonant Frequency on Wave Number.

AT Quartz, $W = .0106\text{ m}$, $L = .02032\text{ m}$,

$F_0 = 8.0\text{ MHz}$, $T = 25^\circ\text{C}$, $n_1 = 0$

W is the plate width (x_1 direction), L is the plate length (x_3 direction), F_0 is the resonant frequency of the thickness twist mode for infinite lateral dimensions ($n_1 = 0$, $n_3 = 0$), and T is the temperature in degrees Centigrade.

DEPENDENCE OF RESONANT FREQUENCY ON WAVE NUMBER

AT QUARTZ, $W = .0106\text{m}$, $L = .02032\text{m}$, $F_0 = 8.0\text{MHz}$, $T = 25^\circ\text{C}$, $n_1 = 10$

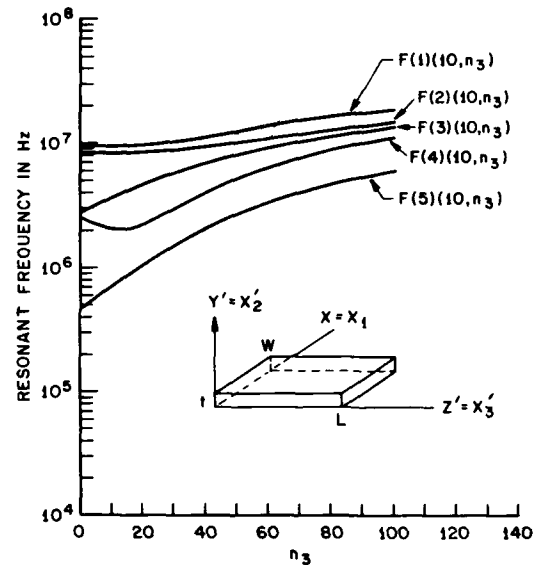


Figure 3

Dependence of Resonant Frequency on Wave Number.

AT Quartz, $W = .0106\text{ m}$, $L = .02032\text{ m}$,

$F_0 = 8.0\text{ MHz}$, $T = 25^\circ\text{C}$, $n_1 = 10$

DEPENDENCE OF RESONANT FREQUENCY ON WAVE NUMBER

AT QUARTZ, $W = .0106$ m, $L = .02032$ m, $F_0 = 8.0$ MHz, $T = 25^\circ\text{C}$

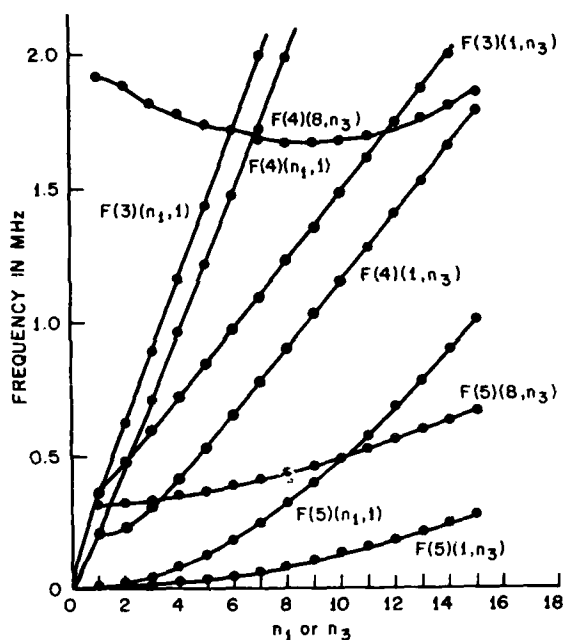


Figure 4
Dependence of Resonant Frequency on Wave Number.
AT Quartz, $W = .0106$ m, $L = .02032$ m,
 $F_0 = 8.0$ MHz, $T = 25^\circ\text{C}$. Solid points are calculated
resonant frequencies.

DEPENDENCE OF n_1 ON n_3 FOR MODES WITH RESONANT FREQUENCIES BETWEEN 7.9 AND 8.1 MHz

AT QUARTZ, $W = .0106$ m, $L = .02032$ m, $F_0 = 8.0$ MHz, $T = 25^\circ\text{C}$

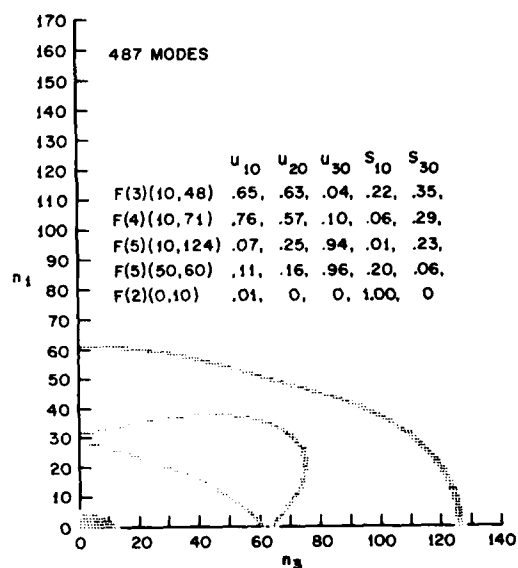


Figure 5

Dependence of n_1 on n_3 for Modes with Resonant Frequencies
between 7.9 and 8.1 MHz
AT Quartz, $W = .0106$ m, $L = .02032$ m,
 $F_0 = 8.0$ MHz, $T = 20^\circ\text{C}$.
The contribution of the five elementary displacement components
to the eigenvectors for five modes are shown.

AN ANALYSIS OF OVERTONE MODES IN CONTOURED CRYSTAL RESONATORS

H.F. Tiersten
Department of Mechanical Engineering,
Aeronautical Engineering & Mechanics
Rensselaer Polytechnic Institute
Troy, New York 12181

R.C. Smythe
Piezo Technology Incorporated
Orlando, Florida 32804

Abstract

A previous treatment of overtone modes in trapped energy resonators is extended to the case of plates with slowly varying thickness. The resulting single scalar equation is applied in the analysis of plano-convex contoured quartz crystal resonators. Both thickness-shear and thickness-twist modal directions are considered. The analysis holds for the fundamental and anharmonic overtones of the fundamental and each harmonic overtone thickness mode. The influence of piezoelectric stiffening, electrode mass loading and electrical shorting is included in the analysis. Since in a contoured resonator the influence of the contouring on the trapping far exceeds the influence of the electrode, the edge of the electrode is ignored in the determination of the mode shape and attendant resonant frequency. Similarly, since the contouring causes the mode to be highly trapped in the vicinity of the center of the plate and, hence, not to extend to the edge, the boundary conditions at the edge of the quartz plate are ignored in the analysis. No adjustable parameters are required in the theory. Although the basic piezoelectric differential equation employed here is quite different from the ones that have been employed in similar applications heretofore, the analysis accounting for the contouring has appeared in the literature. It is shown that calculations based on the analysis agree extremely well with experimental results obtained with contoured AT-cut quartz crystal resonators.

1. Introduction

In a recent investigation^{1,2} the three-dimensional equations of linear piezoelectricity, with the aid of certain simplifying assumptions, were applied in the analysis of rotated Y-cut quartz trapped energy resonators with rectangular electrodes operating in overtones of coupled thickness-shear and thickness-twist vibrations. In this paper the previous analysis is extended to the case of plates with slowly varying thickness. The asymptotic dispersion relation for small wavenumbers along the electroded flat plate obtained in the recent analysis^{1,2} of trapped energy resonators enables us to construct the single scalar differential equation of coupled thickness-shear and thickness-twist, which holds for plates with slowly varying thickness. The simplifying assumptions of small piezoelectric coupling, one-dimensional (thickness) dependence of all electrical variables and the neglect of certain small unimportant elastic constants employed in the earlier^{1,2} analysis naturally are employed here also. The influence of piezoelectric stiffening, electrode mass loading and electrical shorting is included nonetheless. Since the influence of the contouring on the trapping is much greater than the influence of the electrode, the edge of the electrode is ignored in the determination of the eigensolution. Since the mode is assumed to be highly trapped in the vicinity of the center of the contoured plate, the boundary conditions at the edge of the plate are ignored in the analysis. The procedure for obtaining the eigensolution follows that of Wilson³, who treated the purely elastic case

using an incorrect equation which had been used earlier by others^{4,5}

The resulting inhomogeneous single scalar equation is applied in the analysis of the forced vibrations of a plano-convex contoured AT-cut quartz crystal resonator and a lumped parameter representation of the admittance is obtained. Calculations based on the analysis are in excellent agreement with the results of experiments on contoured AT-cut quartz crystal resonators.

2. Basic Equations

A schematic diagram of a plano-convex quartz crystal resonator is shown in Fig.1. Since the thickness h is a slowly varying function of x_1 and x_2 , it is appropriate to first consider the solution for the electroded flat plate and then generalize it to be applicable to the contoured plate with slowly varying thickness. Accordingly, we first briefly reproduce the analysis^{1,2} of the electroded flat plate. On the basis of the simplifying assumptions of small piezoelectric coupling, the neglect of certain relatively small unimportant elastic constants and the fact that in the essentially thickness-shear modes of interest, the wavenumbers in both the x_1 - and x_2 -directions are much smaller than the thickness wavenumber, it has been shown² that to second order in small quantities the differential equations that remain to be satisfied take the form

$$\begin{aligned} c_{11}^{u_1,11} + (c_{12} + c_{66})^{u_2,12} + c_{66}^{u_1,22} + c_{55}^{u_1,33} + \\ e_{26}^{u_1,22} = \rho \ddot{u}_1, \\ (c_{66} + c_{12})^{u_1,12} + c_{66}^{u_2,11} + c_{22}^{u_2,22} = \rho \ddot{u}_2, \\ e_{26}^{u_1,22} - e_{22}^{u_2,22} = 0, \end{aligned} \quad (2.1)$$

and we note that the notation is defined in Ref.2. We further note that for the circumstances outlined it has been shown² that u_2 is two orders of magnitude smaller than u_1 and, hence, negligible to the order of approximation being obtained. To the same order of approximation the pertinent constitutive equations take the form²

$$\begin{aligned} T_{21} &= c_{66}(u_{1,2} + u_{2,1}) + e_{26}\varphi, \\ T_{22} &= c_{12}^{u_1,1} + c_{22}^{u_2,2}, \quad T_{11} = c_{11}^{u_1,1} + c_{12}^{u_2,2}, \\ T_{31} &= c_{55}^{u_1,3} + e_{25}\varphi, \quad D_2 = e_{26}^{u_1,2} - e_{22}\varphi, \end{aligned} \quad (2.2)$$

and the boundary conditions that remain to be satisfied on the electroded major surfaces of the flat plate take the form

$$\begin{aligned} T_{21} &= \mp 2\rho' h' \dot{u}_1, \quad T_{22} = \mp 2\rho' h' \dot{u}_2, \quad \varphi = \pm (V/2)e^{i\omega t} \\ &\text{at } x_2 = \mp h, \end{aligned} \quad (2.3)$$

where ρ' is the mass density of the electrode plating.

It has been shown² that in order to remove the inhomogeneous driving term from the boundary condition (2.3)₃ and place it in the differential equation (2.1)₁, we take

$$u_1 = \hat{u}_1 + Kx_2 e^{i\omega t}, \quad u_2 = \hat{u}_2, \\ \varphi = \frac{Vx_2}{2h} e^{i\omega t} + \frac{e_{26}}{\epsilon_{22}} \hat{u}_1 + Cx_2 e^{i\omega t}, \quad (2.4)$$

where

$$K = -e_{26}V/c_{66}2h, \quad C = -(e_{26}/\epsilon_{22})\hat{u}_1(h)/h. \quad (2.5)$$

Then the substitution of (2.4) and (2.5) into (2.1) and (2.3) yields

$$c_{11}\hat{u}_{1,11} + (c_{12} + c_{66})\hat{u}_{2,12} + \bar{c}_{66}\hat{u}_{1,22} + c_{55}\hat{u}_{1,33} = \\ \rho\ddot{\hat{u}}_1 - \rho\omega^2 Kx_2 e^{i\omega t}, \\ c_{66}\hat{u}_{2,11} + (c_{12} + c_{66})\hat{u}_{1,12} + c_{22}\hat{u}_{2,22} = \rho\ddot{\hat{u}}_2, \quad (2.6)$$

$$c_{66}\hat{u}_{2,1} + \bar{c}_{66}\hat{u}_{1,2} \mp (e_{26}/\epsilon_{22}h)\hat{u}_1 = \mp 2\rho'h'\ddot{\hat{u}}_1 \text{ at } x_2 = \pm h, \\ c_{22}\hat{u}_{2,2} + c_{12}\hat{u}_{1,1} = \mp 2\rho'h'\ddot{\hat{u}}_2 \text{ at } x_2 = \pm h, \quad (2.7)$$

where

$$\bar{c}_{66} = c_{66} + e_{26}^2/\epsilon_{22}, \quad (2.8)$$

and (2.1)₃ and (2.3)₃ are satisfied by the forms chosen. Equations (2.6) and (2.7) constitute a system of linear inhomogeneous differential equations with homogeneous boundary conditions on the major surfaces of the electroded flat plate. It has been shown that an asymptotic eigensolution for plate waves valid to second order in the small wavenumbers ξ and ν along the electroded flat plate can be written in the form

$$\hat{u}_1 = (B_1^{(1)} \sin \eta_1 x_2 + B_1^{(2)} \sin \kappa \eta_1 x_2) \cos \xi x_1 \cos \nu x_3 e^{i\omega t}, \\ \hat{u}_2 = (B_2^{(1)} \cos \eta_1 x_2 + B_2^{(2)} \cos \kappa \eta_1 x_2) \sin \xi x_1 \cos \nu x_3 e^{i\omega t}, \quad (2.9)$$

where

$$\eta_1 = \frac{n\pi}{2h}, \quad \kappa = \sqrt{\frac{\bar{c}_{66}}{c_{22}}}, \quad B_1^{(1)} = \frac{r\xi B_1^{(1)}}{\eta_1}, \quad B_1^{(2)} = -\frac{r\xi B_2^{(2)}}{\kappa\eta_1}, \\ r = \frac{c_{12} + c_{66}}{c_{66} - c_{22}}, \quad B_2^{(2)} = \frac{(-1)^{\frac{n+1}{2}} (c_{12} + rc_{22}) \xi B_1^{(1)}}{c_{22} \kappa \eta_1 \sin \kappa \eta_1 h}, \quad (2.10)$$

and the dispersion relation

$$M_n \xi^2 + c_{55} \nu^2 + \frac{n^2 \pi^2}{4h^2} \bar{c}_{66} - \rho \tilde{\omega}^2 = 0, \quad (2.11)$$

which is valid to second order in ξ and ν must be satisfied, and we employ the definitions

$$M_n = c_{11} + (c_{12} + c_{66})r + 4 \frac{(x\bar{c}_{66} - c_{66})(c_{22}r + c_{12}) \cot \kappa n\pi/2}{c_{22} n\pi \kappa}, \\ \bar{c}_{66} = \bar{c}_{66} \left(1 - \frac{8\kappa^2}{n^2 \pi^2} - 2\hat{R}\right), \quad k_{26}^2 = \frac{e_{26}^2}{\bar{c}_{66} \epsilon_{22}}, \quad \hat{R} = \frac{2\rho'h'}{\rho h}. \quad (2.12)$$

The method of derivation² of the dispersion relation (2.11) reveals that the inhomogeneous differential equation governing coupled thickness-shear and

thickness-twist vibrations is of the form

$$M_n \frac{\partial^2 \hat{u}_1}{\partial x_1^2} + c_{55} \frac{\partial^2 \hat{u}_1}{\partial x_3^2} - \frac{n^2 \pi^2}{4h^2} \bar{c}_{66} \hat{u}_1 - \rho \ddot{\hat{u}}_1 = \rho \omega^2 \frac{e_{26} V x_2}{c_{66} 2h} e^{i\omega t}, \quad (2.13)$$

where ω is the driving frequency. Clearly, the homogeneous ($V=0$) solutions for the flat plate are consistent with (2.11). It has been shown² that when the approximation holds and Eq. (2.13) may be employed, the approximate edge conditions to be satisfied at a junction between an electroded and unelectroded region of the plate are the continuity of

$$\hat{u}_1 \text{ and } \partial \hat{u}_1 / \partial n, \quad (2.14)$$

where n represents the normal to the junction.

We now generalize Eq. (2.13) for the flat plate to be applicable to a contoured plate with slowly varying thickness simply by permitting h in (2.13) to be a slowly varying function of x_1 and x_3 . In the case of a plano-convex crystal with a spherical radius of curvature R and maximum thickness $2h_0 \ll R$, we have^{3,6}

$$2h = 2h_0 [1 - (x_1^2 + x_3^2)/4Rh_0], \quad (2.15)$$

the substitution of which in (2.13) and expansion to first order in $x_1^2 + x_3^2$ yields

$$M_n \frac{\partial^2 \hat{u}_1}{\partial x_1^2} + c_{55} \frac{\partial^2 \hat{u}_1}{\partial x_3^2} - \frac{n^2 \pi^2 \bar{c}_{66}}{4h_0^2} \left[1 + \frac{(x_1^2 + x_3^2)}{2Rh_0}\right] \hat{u}_1 - \\ \rho \ddot{\hat{u}}_1 = \rho \omega^2 \frac{e_{26} V x_2}{c_{66} 2h} e^{i\omega t}, \quad (2.16)$$

which is the inhomogeneous differential equation for coupled thickness-shear and thickness-twist vibrations of a plano-convex resonator.

3. Contoured Resonator

The problem of a contoured resonator driven into coupled thickness-shear and thickness-twist vibrations by the application of a driving voltage across the electrodes in the steady state may now be treated by finding the steady-state solution of (2.16), which remains bounded and vanishes at infinity. We first seek the eigensolutions of the associated homogeneous problem, i.e., with $V=0$. To this end we take the \hat{u}_1 -displacement field in the form

$$\hat{u}_1(x_1, x_2, x_3, t) = u(x_1, x_3) \sin(n\pi x_2/2h) e^{i\omega t}, \quad (3.1)$$

the substitution of which in (2.16) yields

$$M_n \frac{\partial^2 u}{\partial x_1^2} + c_{55} \frac{\partial^2 u}{\partial x_3^2} - \frac{n^2 \pi^2 \bar{c}_{66}}{4h_0^2} \left[1 + \frac{(x_1^2 + x_3^2)}{2Rh_0}\right] u + \rho \tilde{\omega}^2 u = 0, \quad (3.2)$$

where $\tilde{\omega}$ denotes the eigenfrequency. As a solution of (3.2) we take³

$$u = X(x_1)Z(x_3) \quad (3.3)$$

which satisfies (3.2) provided

$$X'' + (\gamma^2 - \alpha^2 x_1^2)X = 0, \quad Z'' + (\mu^2 - \beta^2 x_3^2)Z = 0, \quad (3.4)$$

where the undetermined separation constants γ and μ satisfy

$$\tilde{\omega}^2 - \frac{n^2 \pi^2}{4h_c^2} \bar{c}_{66} = M_n \gamma^2 + c_{55} \mu^2, \quad (3.5)$$

and

$$\alpha_n^2 = n^2 \pi^2 c_{66}^2 / 8Rh_0^3 M_n, \quad \beta_n^2 = n^2 \pi^2 c_{66}^2 / 8Rh_0^3 c_{55}. \quad (3.6)$$

The only solutions of (3.4) that are bounded for all x_1 and x_3 and vanish at ∞ are the Hermite functions

$$x_{mn} = e^{-\frac{\alpha_n}{2} x_1^2} H_m(\sqrt{\alpha_n} x_1), \quad z_{pn} = e^{-\frac{\beta_n}{2} x_3^2} H_p(\sqrt{\beta_n} x_3), \quad (3.7)$$

where H_m and H_p are Hermite polynomials and

$$\alpha_n = \alpha_n (1 + 2m), \quad \beta_n = \beta_n (1 + 2p), \quad m, p = 0, 2, 4, \dots, \quad (3.8)$$

which are determined from the condition that the series for H_m and H_p terminate and they be polynomials. For a given value of n , m and p , (3.8), with (3.6), determines the values of the separation constants γ_{mn} and μ_{pn} , which, from (3.5), yield the eigenfrequencies $\tilde{\omega}_{nmp}$ in the form

$$\tilde{\omega}_{nmp}^2 = \frac{n^2 \pi^2 c_{66}^2}{4h_0^2} \left[1 + \frac{1}{n\pi} \sqrt{\frac{2h_0}{R}} \left(\sqrt{\frac{M_n}{c_{66}}} (2m+1) + \sqrt{\frac{c_{55}}{c_{66}}} (2p+1) \right) \right]. \quad (3.9)$$

We now write the steady-state solution of (2.16) as a sum of eigensolutions, thus

$$\tilde{u}_1 = e^{i\omega t} \sum_n \sum_m \sum_p H_{nmp} \tilde{u}_{1nmp}, \quad (3.10)$$

where

$$\tilde{u}_{1nmp} = \tilde{u}_{1nmp} e^{i\tilde{\omega}_{nmp} t}, \quad \tilde{u}_{1nmp} = \sin \frac{n\pi x_2}{2h} u_{nmp}, \quad (3.11)$$

and

$$u_{nmp} = e^{-\frac{\alpha_n}{2} x_1^2} H_m(\sqrt{\alpha_n} x_1) e^{-\frac{\beta_n}{2} x_3^2} H_p(\sqrt{\beta_n} x_3). \quad (3.12)$$

Substituting from (3.10) and (3.11) into (2.16) and employing (3.2) for each nmp-eigensolution, we obtain

$$\sum_n \sum_m \sum_p H_{nmp} (\omega^2 - \tilde{\omega}_{nmp}^2) \sin \frac{n\pi x_2}{2h} u_{nmp} = \frac{e_{26}^2 v x_2}{c_{66}^2 2h}, \quad (3.13)$$

and with the aid of the orthogonality of the solution functions in (3.11)₂, we find

$$H_{nmp} = \frac{(-1)^{\frac{n-1}{2}} e_{26}^2 4h_0^2 v}{c_{66}^2 n^2 \pi^2 L_{nmp} [1 - (\tilde{\omega}_{nmp}^2 / \omega^2)]}, \quad (3.14)$$

where

$$L_{nmp} = \frac{m! 2^m m! 2^p p!}{\sqrt{\alpha_n} \sqrt{\beta_n}}, \quad J_{nmp} = 4F_{1nm} F_{3np}, \quad (3.15)$$

and if, as is reasonable in the case of the contoured resonator because the mode shape is sharply confined to the center, we replace the circular electrode by the circumscribed square with lengths $2l_1 = 2l_3$, we have

$$F_{1nm} = \int_0^{l_1} e^{-\frac{\alpha_n}{2} x_1^2} H_m(\sqrt{\alpha_n} x_1) dx_1, \quad F_{3np} = \int_0^{l_3} e^{-\frac{\beta_n}{2} x_3^2} H_p(\sqrt{\beta_n} x_3) dx_3. \quad (3.16)$$

In the vicinity of a resonance, say the NMPth, one term in the sum in (3.10) dominates and the others are negligible. Thus in the vicinity of said resonance the steady-state solution may be written

$$u_1 = H_{NMP} \sin \frac{n\pi x_2}{2h} u_{NMP} e^{i\omega t} - \frac{e_{26} v x_2}{c_{66}^2 2h} e^{i\omega t}, \quad \varphi = \frac{v x_2}{2h} e^{i\omega t} + \frac{e_{26}}{\epsilon_{22}} H_{NMP} u_{NMP} \left(\sin \frac{n\pi x_2}{2h} - (-1)^{\frac{n-1}{2}} \frac{x_2}{h} \right) e^{i\omega t} \quad (3.17)$$

where, as usual, ω_{NMP} in (3.17) is to be replaced by

$$\hat{\omega}_{NMP} = \omega_{NMP} + i\omega_{NMP}/2Q_{NMP}, \quad (3.18)$$

in which Q_{NMP} is the unloaded quality factor of the contoured resonator in the NMPth mode. The admittance Y_{NMP} of this rotated Y-cut contoured resonator operating in the NMPth mode is obtained by substituting from (3.17) into (2.2)₅, which is then substituted into

$$I = - \int_A \dot{b}_2 dx_1 dx_3, \quad (3.19)$$

with the result⁶

$$Y_{NMP} = \frac{I}{V} = \frac{i\omega \epsilon_{22}}{2h} (1 + k_{26}^2) \hat{A}_e + \frac{i\omega \epsilon_{22} k_{26}^2 4J_{NMP}^2}{[(\hat{\omega}_{NMP}^2 / \omega^2) - 1] N^2 \pi^2 L_{NMP}}, \quad (3.20)$$

where

$$k_{26}^2 = e_{26}^2 / c_{66} \epsilon_{22}, \quad \hat{A}_e = A_e (1 + l_1^2 / 8Rh_0), \quad (3.21)$$

and in obtaining the second term in (3.20) we have again replaced the circular electrode by the circumscribed square with lengths $2l_1 = 2l_3$ in order to perform the integrations. The quantities C_0 and C_{NMP} defined by

$$C_0 = \frac{\hat{A}_e \epsilon_{22} (1 + k_{26}^2)}{2h}, \quad C_{NMP} = \frac{4\epsilon_{22} k_{26}^2 J_{NMP}^2}{N^2 \pi^2 L_{NMP}}, \quad (3.22)$$

are called the static and motional capacitances, respectively. The integrals appearing in (3.16), which appear prominently in (3.22)₂, have been evaluated for $M = P = 0$ and take the form

$$F_{1NO} = \frac{1}{\sqrt{\alpha_N}} \sqrt{\frac{\pi}{2}} \operatorname{erf} \sqrt{\frac{\alpha_N}{2}} l_1, \quad F_{3NO} = \frac{1}{\sqrt{\beta_N}} \sqrt{\frac{\pi}{2}} \operatorname{erf} \sqrt{\frac{\beta_N}{2}} l_3. \quad (3.23)$$

Equation (3.9) has been employed in the calculation of some resonant frequencies of two plano-convex resonators, which are compared with frequency measurements on the respective resonators. The first resonator has a blank diameter of .550 in., an electrode diameter of .370 in., a radius of curvature R of 5 diopters, which is 106 mm, an electrode mass loading ratio R of 1.864×10^{-3} and a measured thickness $2h_0 = .0271$ in. = .6883 mm. Since we did not have confidence in the accuracy of the measured thickness to the required number of significant figures, we adjusted the thickness in order that the

calculated value of $f(3,0,0)$ agree with the measured value. The adjusted thickness $2h_0$ is .68785 mm. The comparison between the calculated and measured values is given in Table I, which shows that the agreement between theory and experiment is excellent with the exception of $f(1,2,2)$. We believe that strong coupling to flexure, which has been omitted from the theory, exists in the case of the mode for which the calculation does not agree well with the measurements. The measured frequencies in Table I are the average values determined from five units. For this resonator the motional capacitance of the fundamental mode, i.e., C_{100} , was calculated from Eq. (3.22) and compared with the average value determined from the measurement of five units. The results are

$$\text{Calculated } C_{100} = 12.51 \text{ ff}, \text{ Measured } C_{100} = 13.21 \text{ ff.} \quad (3.24)$$

The second resonator has a blank diameter of .590 in., an electrode diameter of .374 in., a radius of curvature R of 2.0 in., which is .0508 m, a measured thickness of gold electrode $2h' = 750 \text{ \AA}$ and a maximum plate thickness $2h_0 = .06502 \text{ in.} = 1.6515 \times 10^{-3} \text{ m}$. In this case we used the measured maximum plate thickness, since the measurement is considered to be accurate to one more significant figure than in the case of the first resonator, and, consequently, no adjustable parameters were employed. The comparison between the calculated and measured values is given in Table II, in which excellent agreement between theory and experiment is indicated again.

Acknowledgements

We wish to thank Drs. T.R. Meeker and A.A. Comparini of Bell Laboratories for providing the measured data associated with Table II.

The work of one of the authors (HFT) was supported in part by the Army Research Office under Grant No. DAAG29-76-G-0176 and the Office of Naval Research under Contract No. N00014-76-C-0368.

References

1. H.F. Tiersten, "Analysis of Trapped Energy Resonators Operating in Overtones of Coupled Thickness-Shear and Thickness-Twist," *Proceedings of the 29th Annual Symposium on Frequency Control*, U.S. Army Electronics Command, Fort Monmouth, New Jersey, 71 (1975).
2. H.F. Tiersten, "Analysis of Trapped Energy Resonators Operating in Overtones of Coupled Thickness-Shear and Thickness-Twist," *J. Acoust. Soc. Am.*, 59, 879 (1976).
3. C.J. Wilson, "Vibration Modes of AT-Cut Convex Quartz Resonators," *J. Phys. D: Appl. Phys.*, 7, 2449 (1974).
4. H.J. McSkimin, in *Quartz Crystals for Electrical Circuits*, edited by R.A. Heising (D. Van Nostrand, New York, 1946), Chap.VII.
5. W.G. Stoddard, "Design Equations for Plano-Convex AT Filter Crystals," *Proceedings of the 17th Annual Symposium on Frequency Control*, U.S. Army Electronics Command, Fort Monmouth, New Jersey, 272 (1963).
6. For more detail see H.F. Tiersten and R.C. Smythe "An Analysis of Contoured Crystal Resonators Operating in Overtones of Coupled Thickness-Shear and Thickness-Twist," to be issued as a technical report, Rensselaer Polytechnic Institute, Troy, N.Y. 12181.
7. P.M. Morse and H. Feshbach, *Methods of Theoretical Physics*, Part I (McGraw-Hill, New York, 1953), Chap.5.

TABLE I

Mode N M P	Calculated Frequency kHz	Measured Frequency kHz
1 0 0	2508.2	2505.5
1 0 2	2684.0	2683.4
1 2 0	2728.6	2727.7
1 2 2	2891.0	2843.2
3 0 0	7325.8	7325.8
3 0 2	7510.9	7514.1
3 2 0	7520.0	7520.1
3 2 2	7700.5	7693.4
5 0 0	12152.6	12154.1
5 0 2	12339.6	12343.0
5 2 0	12366.3	12367.7
5 2 2	12550.1	12532.0

TABLE II

Mode N M P	Calculated Frequency kHz	Measured* Frequency kHz
1 0 0	1097.47	1094.75
1 0 2	1248.72	1245.93
1 2 0	1285.69	1263.23
1 2 2	1416.99	-
3 0 0	3103.25	3103.54
3 0 2	3270.74	3278.93
3 2 0	3278.82	3285.67
3 2 2	3437.77	3430.30
5 0 0	5123.38	5120.14
5 0 2	5294.40	5297.28
5 2 0	5318.61	5319.72
5 2 2	5483.54	5472.17
7 0 0	7127.01	7129.09
7 0 2	7299.73	7292.75
7 2 0	7318.23	7320.40
7 2 2	7486.54	7468.37

* Mode identification is conjectural. Measured data courtesy of T.R. Meeker and A.A. Comparini of Bell Laboratories.

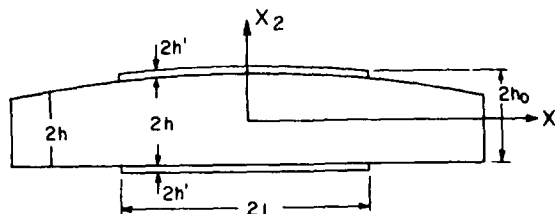


Figure 1 Plano-Convex Resonator

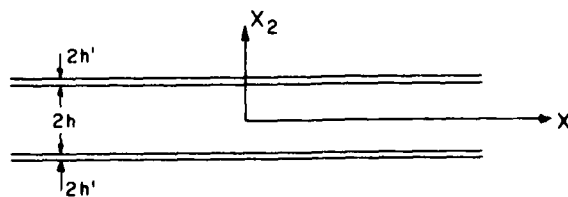


Figure 2 Electroded Flat Plate

4 MHZ AT-CUT STRIP RESONATOR FOR WRIST WATCH

Morio ONOE

Institute of Industrial Science
University of Tokyo
Roppongi, Tokyo, Japan

and

Keinosuke KAMADA, Masanobu OKAZAKI, Fumitake TAJIKA, Naoki MANABE

Nihon Dempa Kogyo Co., Ltd.
Shibuya-ku, Tokyo, Japan

Summary

The use of miniature quartz resonators for wrist watches has increased rapidly in recent years. Most of these resonators are either bar-shaped flexural resonators or tuning fork resonators at 32 kHz. A bar-shaped resonator is easy to manufacture, but difficult to mount in a limited space as to withstand severe shock and vibration. On the other hand, a tuning fork resonator is inherently resistant to shock and vibration, but requires more complicated manufacturing, especially in splitting arms and shaping electrodes. Furthermore, the bar-shaped resonator exhibits a rather large aging and both types of resonators have parabolic frequency temperature curve. Consequently, the accuracy of wrist watches using these types of resonators is on the order of 10 seconds per month.

Therefore, there has been a strong demand for a miniature quartz resonator featuring low aging, good frequency temperature curve, good productivity and high resistance against shock and vibration. One candidate to meet this demand is an AT-cut strip resonator with tilted edges previously reported.(1) The tilting of the edge planes off normal relative to the major surface eliminates coupling between the desired thickness shear mode and the undesired face shear modes which would otherwise cause strong unwanted responses and an undesirable change in the frequency temperature curve when the width dimension is reduced for miniaturization.

Zumsteg and Suda reported on properties of a 4 MHz miniature flat rectangular quartz resonator vibrating in a coupled mode.(2) A resonator elongated in the Z direction vibrates in a coupled mode between the thickness shear mode and flexure mode. In this case, the location of the inflection point of the cubic frequency temperature curve shifts to a higher temperature, as Royer has suggested already.(3) This shift cannot be corrected by adjustment of the cut angle.

This paper describes the development of 4.2 MHz AT-cut strip resonators with tilted edges for wrist watches. A good cubic frequency temperature characteristics are realized by selection of a combination between the cut angle and tilt angle. A low CI value and high quality factor can be realized by beveling and mounting securely at both ends along the length. Therefore high resistance against shock and vibration is realized. Presently the resonator is sealed in a cold-welded flat package, and low frequency aging characteristics are obtained. The resonator is miniature enough for wrist watch application. An SOS C-MOS LSI is used for the oscillator and frequency dividers. Total current consumption is presently 8 μ A at 1.5 V. A wrist watch using this circuit and one silver oxide battery can keep time as accurately

as 1 second per month for one year.

Key Words

AT strip with tilted edges, Beveling, Cold-welded flat package, Cubic frequency temperature curve, Low aging, SOS C-MOS LSI

Introduction

A rotated Y-cut quartz resonator, such as AT or BT, vibrates in a similar thickness twist mode in which the direction of displacement is parallel to the diagonal, X, axis. Fig. 1 (a) shows the thickness twist mode of a large isotropic plate. Edge planes perpendicular to the major surface are traction-free for displacement of the pure thickness twist mode. This isotropic case, a strip with a rectangular cross section has been used to obtain a miniaturized resonator as shown in Fig. 1 (b).

Because of anisotropy, however, edge planes perpendicular to the major surface are no longer traction-free for displacement of the pure thickness twist mode of an infinite quartz plate. Tilting of the edge planes off normal to the major surfaces satisfied the traction-free boundary conditions of edge planes as shown in Fig. 1 (c) and, thus, eliminates the coupling between the desired thickness shear mode and the undesired face shear modes which would otherwise cause strong unwanted responses and undesirable change in the frequency temperature curve when the width dimension is reduced for miniaturization.

A rotated Y-cut strip resonator with its coordinate system is shown in Fig. 2. The length of the plate along the X-axis is large compared with the width and thickness. θ is the cut angle and α is the tilt angle of the edges. Fig. 3 shows the typical shape of a resonator. The length along the X-axis is $2a$, Z-axis is $2c$ and the thickness is $2h$. The range of useful tilt angle is 5 to 10 degrees for AT-cut.

A miniature AT-cut strip resonator with tilted edges has less unwanted responses and good cubic frequency temperature characteristics. This paper discusses the development of a 4.2 MHz strip resonator for wrist watch as one application of this resonator.

Design and Fabrication

The thickness is determined by the resonant frequency, but the contour dimension along the X and Z-axis can be freely chosen. The length along the X-axis is not critical and is chosen as 25 times the thickness. The length along the Z-axis without coupling with the face shear mode is chosen (to be 4 times the thickness). The frequency spectrum around the fundamental thickness twist mode is shown in Fig. 4

as a function of the width versus thickness ratio. The horizontal line is for the thickness twist mode and the slopes are for the overtone of the face shear modes.

The frequency temperature characteristics are controlled by selection of a combination between the cut angle and tilt angle. Fig. 5 shows the change in the effective cut angle at the frequency temperature characteristics as a function of the cut angle at $c/h=4$, $a/h=25$ contour dimension with 5 degrees tilt. The effective cut angle is the cut angle of an infinite plate calculated from the frequency temperature characteristics of a finite plate. A zero temperature coefficient is realized at a cut angle of $35^{\circ}09'$.

The CI value has a close dependence on the length along the X-axis because the leakage of vibration energy also depends on it. Beveling both ends along the X-axis is effective in preventing energy leakage. The changes in the CI value and quality factor achieved by beveling both ends are shown in Fig. 6. The CI value decreases to 1/10 and the quality factor increases eight times after beveling. Fig. 7 shows the picture of a strip resonator beveled both ends.

The resonator is securely mounted at both beveled ends along the length as shown in Fig. 8. This mounting structure strengthens the resistance against shock and vibration without adversely affecting the properties of the resonator, and is suitable for sealing in a flat package. Presently the resonator is sealed in an X-41 or X-31 type cold-welded flat package. Fig. 9 shows the X-41 type package of $11.7 \times 2.6 \times 4.1$ mm. Fig. 10 shows the X-31 package of $12 \times 3.1 \times 3.1$ mm.

Characteristics

As shown in Fig. 11 (a), (b), the resonator has a clean frequency spectrum. The measured frequency responses over a wide range from 1.5 to 6.5 MHz is shown in (a). The first face shear mode appears at nearly 1.6 MHz. The frequency responses over a narrow range from 3.7 to 4.7 MHz is shown in (b).

The frequency temperature characteristics over the -10°C to $+80^{\circ}\text{C}$ range are shown in Fig. 12. The frequency change is less than 10 ppm and is nearly one-fifth as large as that of a tuning fork resonator. For wrist watch applications, the location of the inflection point of the cubic frequency temperature curve should be slightly higher than the room temperature.

Typical equivalent constants of the resonator are given in Table 1. The CI value of 500 is low enough for oscillation with an C-MOS LSI. The maximum frequency change due to temperature change, vibration, shock and aging is shown in Table 2. The resonator is extremely resistant to severe shock and vibration.

The special support method and cold-welded enclosure provide the resonator with a reliable characteristics as shown in Fig. 13. Because of its parallel capacitance C_0 1.8 pF and capacitance ratio of C_0/C_1 of 280, the typical frequency versus load capacitance curve of the resonator has the wide adjustable range as shown in Fig. 14.

Oscillator and Dividers

A low current consumption, high-speed operation circuit is required for watch applications. Furthermore, the circuit must also be stable against changes

in the DC supply voltage. Therefore, an SOS (Silicon on Sapphire) C-MOS LSI is used for the oscillator and frequency dividers as shown in Fig. 15.

The measured frequency change as a function of the DC supply voltage is shown in Fig. 16. The oscillation starting voltage is 1.2 V at a CI of 600. If CI is increased to 1100, the starting voltage increases to 1.3 V. The change in total current consumption as a function of the DC supply voltage is shown in Fig. 17. Total current consumption is 80A at 1.5 V, 600, and increases only slightly even if the CI is increased to 1100. The oscillation frequency is adjustable by controlling the parallel load capacitance in the oscillation circuit. The frequency is adjustable over the 120 ppm by changing the parallel load capacitance between 0.5 and 3 pF as illustrated in Fig. 18. A wrist watch using this circuit and one silver oxide battery can keep time at an accuracy of 1 second per month for one year.

Miniaturization

4.2 MHz quartz resonators more miniature than the X-41 and X-31 type will be demanded in the future. Such resonators are being developed based on the theory and experimental results of the X-41 and X-31 type resonators. The dimensions of the resonator are approximately 6.5 mm long, 2 mm wide, 0.4 mm thick. Fig. 19 shows this superminiature strip resonator beveled both ends. These reduced dimensions have no effect on the CI value. Good cubic frequency temperature characteristics can be maintained by suitable selection of a combination between cut angle and tilt angle. Miniaturization has no adverse effect on the other characteristics either. Presently the resonator is sealed in a cold-welded $8 \times 2 \times 3$ mm flat package.

Conclusion

A 4.2 MHz AT-cut strip resonator for wrist watches has been developed by utilizing the principle of tilting the edges. Good cubic frequency temperature characteristics have been achieved by suitable selection of the cut angle and tilt angle. The resonator is securely mounted at both beveled ends along the length to realize a clean frequency spectrum and high resistance against shock and vibration. Presently the resonator is sealed in a cold-welded flat package and a low aging ratio is obtained.

Typical characteristics are as follows:

Nominal frequency	4.194304 MHz
Frequency tolerance at 25°C	+20 ppm
Maximum drive level	50 μW
Maximum series resistance	100 ohms
Frequency vs. temperature variation over the -10°C to $+80^{\circ}\text{C}$	10 ppm max.
Aging ratio	First month First year
	1 ppm max. 2 ppm max.
Maximum frequency change due to vibration under MIL-STD-202E method 201A conditions.	2 ppm max.
Maximum frequency change due to shock under MIL-STD-202E method 207A conditions	2 ppm max.
Parallel capacitance	1.8 pF
Capacitance ratio	280

A wrist watch using an SOS oscillator and divider circuit and one silver oxide battery can keep time at an accuracy of 1 second per month for one year.

References

- (1) M. Onoe and M. Okazaki:
"Miniature AT-cut strip resonator with tilted edges"
Proc. 29th Freq. Control Symp.,
pp.42-48, 1975
 - (2) A.E. Zumsteg and P. Suda:
"Properties of a 4MHz miniature flat rectangular quartz resonator vibrating in a coupled mode"
Proc. 30th Freq. Control Symp.,
pp.196-201, 1976
 - (3) J.J. Royer:
"Rectangular AT-cut resonators"
Proc. 27th Freq. Control Symp.,
pp.30-34, 1973
- * Nihon Dempa Kogyo Co. Ltd.
1-21 Nishihara, Shibuya-ku
Tokyo 151, Japan

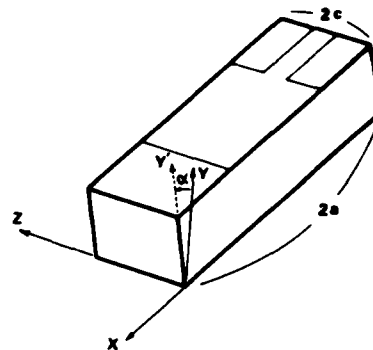


Fig. 3 Shape of a strip resonator with tilted edges.

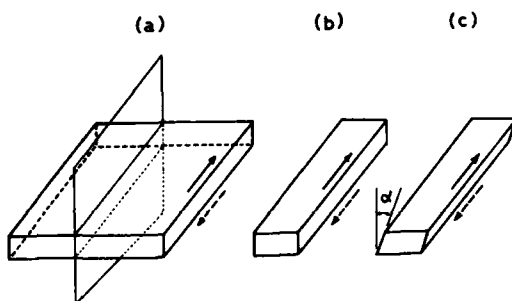


Fig. 1 Principle of miniaturization of thickness twist resonator.

- (a) Traction-free plane in a large resonator.
- (b) Strip resonator with rectangular cross-section.
(isotropic case)
- (c) Strip resonator with parallel piped cross-section.
(anisotropic case)

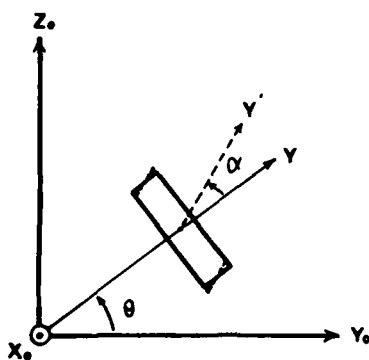


Fig. 2 Strip resonator with it's coordinate system.

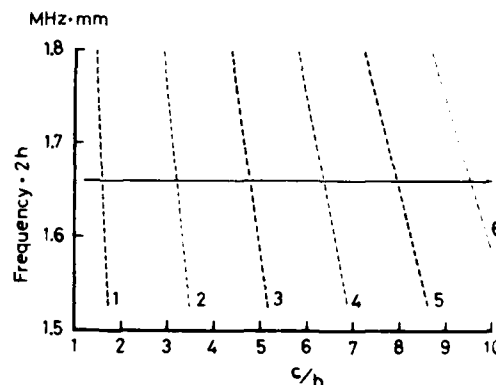


Fig. 4 Frequency spectrum of the exact solution as a function of the width versus thickness ratio.

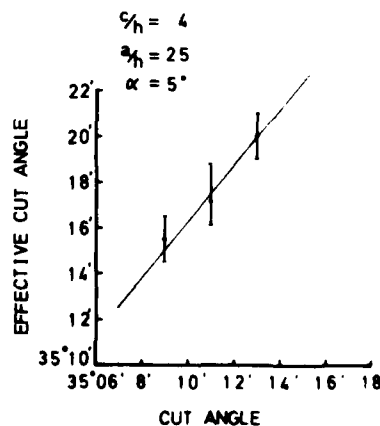


Fig. 5 Change of the effective cut angle at frequency temperature characteristics as a function of the cut angle.

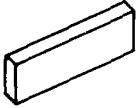
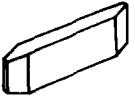
$c_h = 4$ $a_h = 25$		
C.I.	600 ~ 1000 Ω	30 ~ 150 Ω
Q	0.5 ~ 2.0 $\times 10^4$	4.0 ~ 15.0 $\times 10^4$

Fig. 6 Change in the CI value and quality factor achieved by beveling both ends.

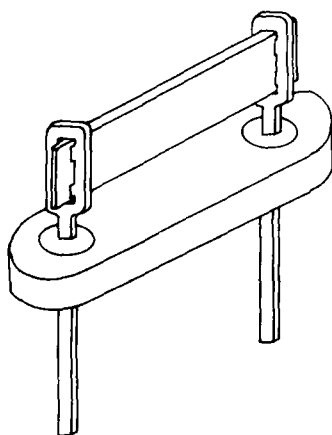


Fig. 8 Mounting structure.

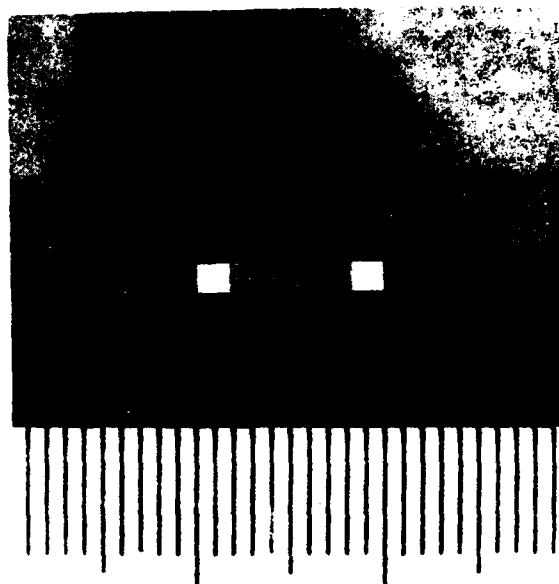


Fig. 7 Picture of a strip resonator beveled both ends.

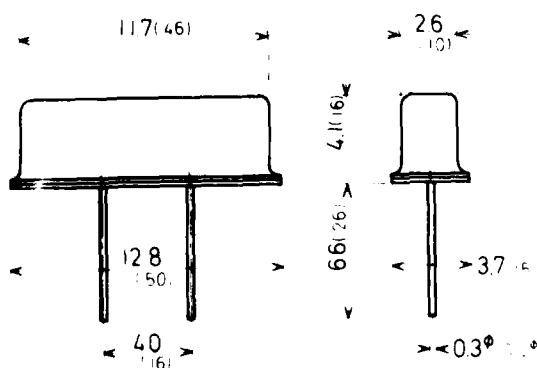


Fig. 9 X-41 type cold-welded flat package.

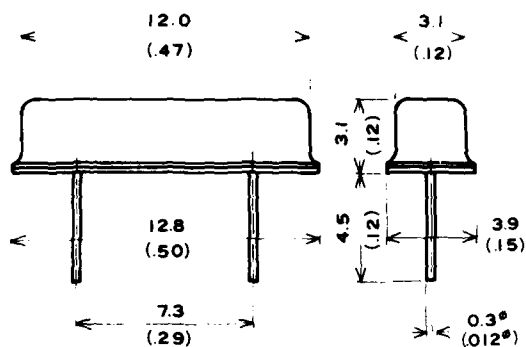


Fig. 10 X-31 type cold-welded flat package.

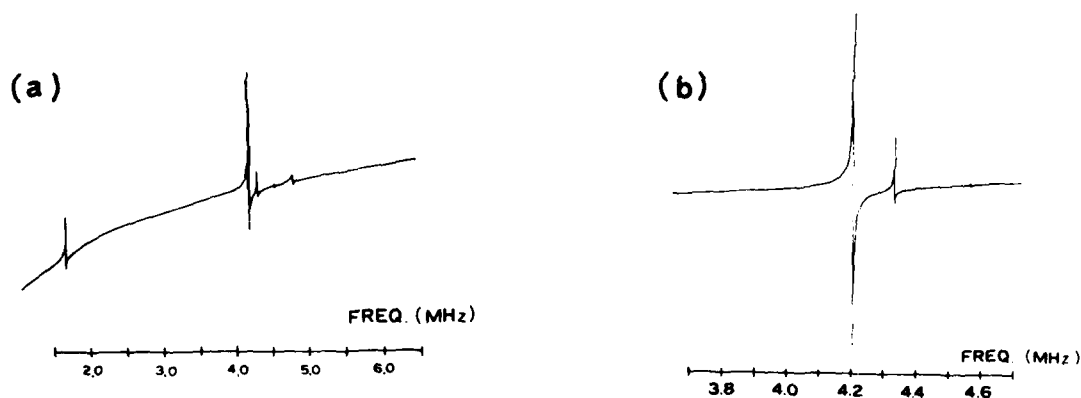


Fig. 11 Measured frequency responses.
 (a) Wide range from 1.5 to 6.5 MHz.
 (b) Narrow range from 3.7 to 4.7 MHz.

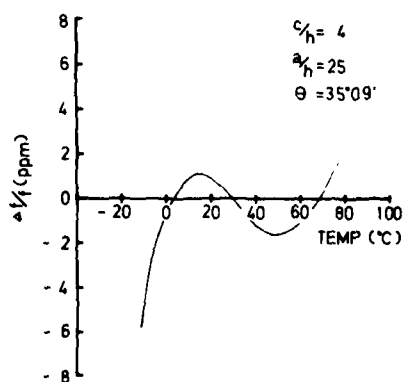


Fig. 12 Measured frequency temperature characteristics over the -10°C to $+80^\circ\text{C}$ range.

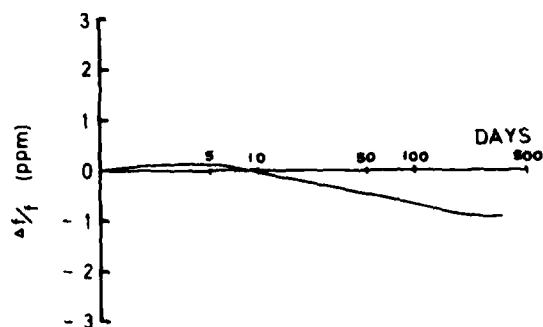


Fig. 13 Frequency aging characteristics of the resonator.

	TYP. VALUE	UNIT
Fosc.	4.194304	MHz
Rs	50	Ω
Co	1.8	pF
C1	0.0064	pF
Co/C1	280	
L1	230	mH
Q	12×10^5	

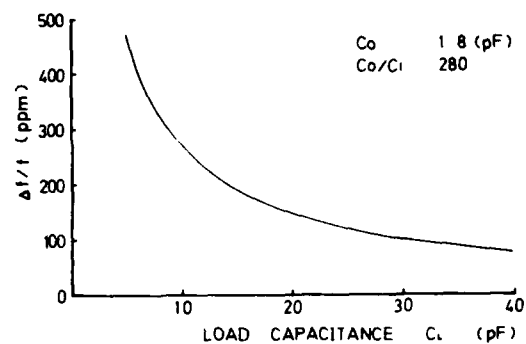


Fig. 14 Frequency vs. load capacitance characteristics of the resonator.

Table 1 Typical equivalent constants of the resonator.

	CONDITION	Max. FREQ CHANGE
Freq. Temp. Characteristics	-10°C - +80°C	20 ppm
Vibration	MIL-STD-202E METHOD 201A	2 ppm
Shock	50 cm Random Dropping	2 ppm
Aging	First 30 Days	1 ppm
	First Year	2 ppm

Table 2 Maximum frequency change due to temperature change, vibration, shock and aging.

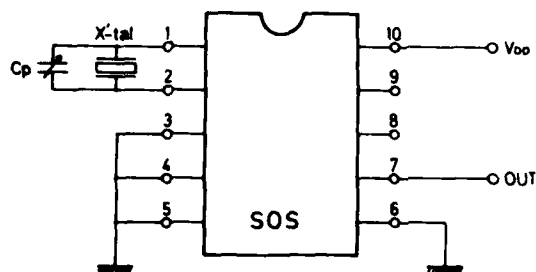


Fig. 15 Oscillator and frequency dividers circuit using SOS CMOS LSI.

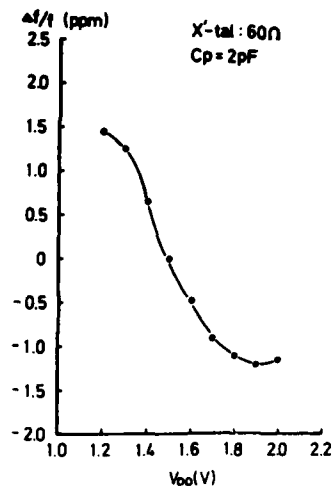


Fig. 16 Measured frequency change as a function of the DC supply voltage.

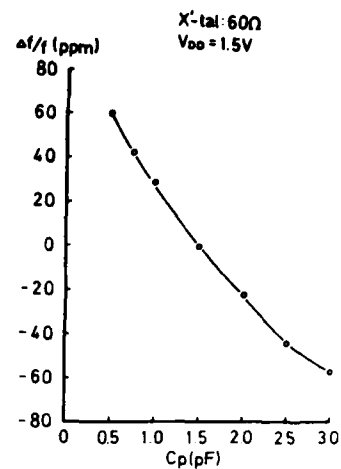


Fig. 18 Measured frequency change as a function of the parallel load capacitance.

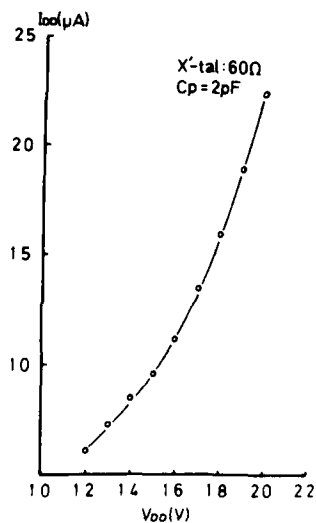


Fig. 17 Total current consumption as a function of the DC supply voltage.

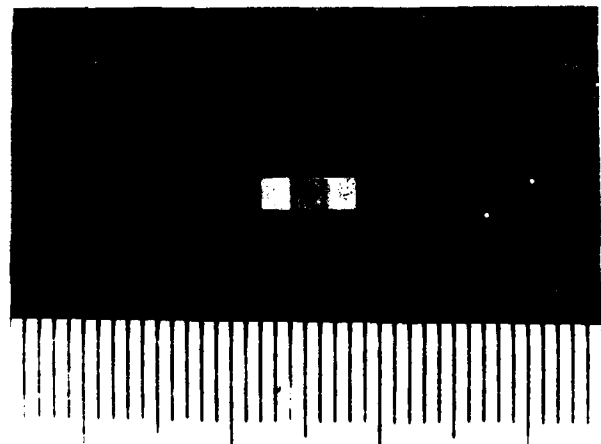


Fig. 19 Picture of a superminiature strip resonator beveled both ends.

DT - CUT TORSIONAL RESONATORS

by

J. Hermann
Centre Electronique Horloger S.A.
Neuchâtel, Switzerland

Summary

The properties of quartz bars vibrating in torsion are studied, the analysis being restricted to resonators with a rectangular cross section and the length oriented along the X-axis.

Several relations are developed for the determination of the following quantities: frequency, stress distribution, piezoelectric coupling factor, as well as static and motional capacitances.

Furthermore, curves are presented which give the first and second-order temperature coefficient of frequency as a function of the angle of cut (X-rotation) and of the dimensional ratio of the cross section. These curves show that for two angles of cut, corresponding approximately to the DT- and CT-cuts known for surface shear resonators, the first-order temperature coefficient, which is then zero, is nearly independent of the dimensional ratio of the cross section. The second-order temperature coefficient of the DT-cut, which lies in the range -15 to $-20 \cdot 10^{-9}/^{\circ}\text{C}^2$, has only half the value usually obtained with flexural and length-extensional resonators.

It is also shown that two torsional bars can be coupled to form a torsional tuning-fork. The properties of such resonators are very similar to those of torsional bars.

Introduction

Few papers have been devoted to quartz resonators with low first-order temperature coefficients vibrating in torsion. Giebe and Blechschmidt¹ have studied the temperature behavior of quartz bars having a rectangular cross section, with edges oriented along the Y- and Z-axes, and the length oriented along the X-axis. They have shown that a zero first-order temperature coefficient of frequency can be obtained by a suitable choice of the dimensional ratio of the cross section. Vasin and Pozdnjakov^{2,3} have described a similar resonator, whose cross section has been rotated by $\pm 45^{\circ}$ about the X-axis, in order to increase the piezoelectric coupling. The drawback of both resonators is a very strong dependence of the first-order temperature coefficient upon the dimensional ratio of the cross section, which calls for very stringent dimensional tolerances.

The analysis presented in this paper yields a quite accurate frequency equation, which allows the determination of the temperature behavior of quartz bars up to the second-order temperature coefficient.

This study leads to a DT-cut torsional resonator, whose zero first-order temperature coefficient is very nearly independent of the dimensional ratio of the cross section.

The analysis has been restricted to bars whose length is oriented along the X-axis for the following reasons:

- (i) There is neither elastic nor piezoelectric coupling to flexural modes.
- (ii) The highest possible piezoelectric coupling to torsion is obtained.
- (iii) A single rotation is sufficient to annul the first-order temperature coefficient of frequency.

Torsion-function and torsional rigidity

The torsion problem has been first studied by Saint-Venant, and later by Love⁴. Their analysis is based on the use of a so-called "torsion-function", which turns out to be proportional to the out-of-plane motion of the cross section.

The length of the bar is oriented along the x-axis, and its cross section lies in the (y, z) plane. Two quantities are introduced:

$$\begin{aligned}\phi(y, z) &: \text{torsion-function} \quad [\text{m}^2] \\ \tau(x) &: \text{torsion angle per unit length} \quad \left[\frac{\text{rad}}{\text{m}} \right]\end{aligned}$$

The displacements are then given by the following relations:

$$\begin{aligned}u &= \tau \phi \\ v &= \tau x z \\ w &= -\tau x y\end{aligned}$$

Consequently, the only non-zero strains are S_5 and S_6 . Moreover, due to the fact that X is a digonal axis, T_5 and T_6 are the only non-zero stresses. This also holds in the resonant case, providing that the cross-sectional dimensions of the bar are small compared to its length.

Love has shown that the shear stresses across a rectangular cross section are equivalent to a couple of moment $C\tau$, with:

$$C = \iint_{\text{cross section}} \left[c_{55} y \left(y - \frac{\partial \phi}{\partial z} \right) + c_{66} z \left(z + \frac{\partial \phi}{\partial y} \right) + c_{56} \left(z \frac{\partial \phi}{\partial z} - y \frac{\partial \phi}{\partial y} \right) \right] dy dz$$

The determination of C , which is known as the torsional rigidity, is greatly simplified in the case $c'_{56} = 0$. This assumption, which leads to small inaccuracies that will be discussed later, has been made throughout this paper. It implies that:

$$s'_{56} = 0 \quad \text{and} \quad s'_{55} c'_{55} = s'_{66} c'_{66} = 1$$

But in fact:

$$c'_{56} \neq 0 \neq s'_{56} \quad \text{and} \quad s'_{55} c'_{55} = s'_{66} c'_{66} \geq 1$$

which means that the simplifying assumption leads to different consequences depending whether the various relations are expressed in terms of elastic stiffnesses, or in terms of elastic compliances. Considerations based upon Hamilton's principle lead to expect the latter choice to yield much more accurate values. This is supported by comparison between theoretical and experimental frequencies. The following substitutions have therefore been made in Love's relations:

$$c'_{55} \rightarrow 1/s'_{55} \quad c'_{66} \rightarrow 1/s'_{66}$$

Fig. 1 shows the cross section of the bar, and defines the dimensionless quantities α , β and z .

The torsion-function and the torsional rigidity are easily obtained from Love's relations.

$$\phi = -yz + \frac{2^5}{\pi^3} \frac{ab}{z} \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^3} \frac{\sinh \frac{(2n+1)\pi\alpha z}{2} \sin \frac{(2n+1)\pi\beta}{2}}{\cosh \frac{(2n+1)\pi\alpha z}{2}}$$

$$C = \frac{1}{s'_{55}} \frac{16ab^3}{3} \left[1 - \frac{192}{\pi^5 z} \sum_{n=0}^{\infty} \frac{\tanh \frac{(2n+1)\pi\alpha z}{2}}{(2n+1)^5} \right]$$

Stress distribution

The relevant stresses are derived from the torsion-function:

$$T_5 = \frac{\tau}{s'_{55}} \left(\frac{\partial \phi}{\partial z} - y \right)$$

$$T_6 = \frac{\tau}{s'_{66}} \left(\frac{\partial \phi}{\partial y} + z \right)$$

so that

$$T_5 = \frac{b\tau}{s'_{55}} \left[\left(\frac{4}{\pi} \right)^2 \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^2} \frac{\cosh \frac{(2n+1)\pi\alpha z}{2} \sin \frac{(2n+1)\pi\beta}{2}}{\cosh \frac{(2n+1)\pi\alpha z}{2}} - 2\beta \right]$$

$$T_6 = \frac{b\tau}{\sqrt{s'_{55}s'_{66}}} \left(\frac{4}{\pi} \right)^2 \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^2} \frac{\sinh \frac{(2n+1)\pi\alpha z}{2} \cos \frac{(2n+1)\pi\beta}{2}}{\cosh \frac{(2n+1)\pi\alpha z}{2}}$$

Each stress is symmetrical with respect to the axis of the cross section normal to its own axis. Its value is zero along two faces, and reaches a maximum in the center of the other two faces. This is illustrated in fig. 2.

It follows from the above equations that, for a markedly rectangular cross section, the main contribution to the elastic energy is made by the shear stress whose axis is normal to the larger faces. But even in this case, the ratio $T_6 \max / T_5 \max$ is not very different from unity. This is shown in fig. 3, which presents this quantity for various cross-sectional ratios as a function of the angle of cut φ (X-rotation).

Fig. 4 and fig. 5 represent both stresses, in normalized form, for $b/a = 1/4$ and $\varphi = -53^\circ$ (DT-cut). It is apparent that T_5 varies almost linearly with the thickness, whereas T_6 presents a kind of "skin-effect".

Frequency equation

It has been shown⁵ that the fundamental torsional frequency of a rod is given by the relation:

$$f = \frac{1}{2L} \sqrt{\frac{C}{\rho I}}$$

where:

L : length of the rod

C : torsional rigidity

ρ : density

$I = \iint_{\text{cross section}} (y^2 + z^2) dy dz$: moment of inertia of the cross section

The frequency may then be expressed by the following fast converging series:

$$f = \frac{1}{L \sqrt{\rho s'_{55}}} \frac{\left(\frac{b}{a} \right)}{\sqrt{1 + \left(\frac{b}{a} \right)^2}} \sqrt{1 - \frac{192}{\pi^5 z} \sum_{n=0}^{\infty} \frac{\tanh \frac{(2n+1)\pi\alpha z}{2}}{(2n+1)^5}}$$

The frequency constant fL relative to three different angles of cut has been plotted in fig. 6 as a function of the dimensional ratio. The variation of this ratio has been restricted to the range $0 \leq b/a \leq 1$, so that the Y' -axis is always normal to the larger faces of the bar. The curve corresponding to the angle $\varphi = -58^\circ$, respectively $\varphi = 32^\circ$, has been selected because it represents nearly an envelope of the set.

The frequency equation has been checked against experimental data from several authors:

- Giebe and Blechschmidt¹: $0.25 \leq b/a \leq 1$, $\varphi = 0^\circ$ and $\varphi = 90^\circ$
- Pozdnjakov³: $0.8 \leq b/a \leq 1$, $\varphi = \pm 45^\circ$
- CEH: $b/a = 0.25$, $\varphi = -53^\circ$

In all cases, the discrepancies are found to be smaller than 1%.

Temperature behavior

The temperature behavior of torsional bars may be determined by means of the frequency equation.

To avoid lengthy analytical derivations, a numerical method is used, which calls for the frequency to be computed at three different temperatures, conveniently chosen symmetrical with respect to 25°C. The temperature coefficients of the various constants involved are taken into account up to the third-order, and the first- and second-order temperature coefficients of frequency are then easily obtained.

The computations have been performed using values given by Bechmann and al.⁶ at constant electric field.

Fig. 7 presents the first-order temperature coefficient as a function of the b/a ratio, for various angles of cut. This figure shows that there exist two angles, approximately equal to -52° and 36°, for which a zero first-order temperature coefficient is obtained for resonators having low b/a ratios. Moreover, these two cuts are characterized by a small sensitivity of the temperature coefficient with respect to the dimensional ratio.

By comparison, the torsional bars studied by Giebe and Blechschmidt¹ ($\Psi = 0^\circ$), and by Vasin and Pozdnjakov^{2,3} ($\Psi = \pm 45^\circ$), exhibit much larger sensitivities, which make these resonators unsuitable for stringent temperature behavior tolerances.

It is interesting to note that the two angles mentioned above (-52°/36°) correspond nearly to the DT- and CT-cuts known for surface shear resonators. This is not surprising, since only a single shear stress contributes to the elastic energy of resonators having markedly rectangular cross sections. Their frequency equation is then similar to that of surface shear resonators.

Fig. 8 shows the second-order temperature coefficient of frequency, whereas fig. 9 gives the first-order temperature coefficient, at an enlarged scale, for DT- and CT-cuts. Fig. 10 presents the corresponding curves for the second-order temperature coefficient, which is seen to be lower for the DT-cut.

Comparisons have been made between computed and measured temperature coefficients for nearly square cross sections, with $\Psi = 0^\circ$ ¹ and $\Psi = \pm 45^\circ$ ³, as well as for markedly rectangular cross sections with $\Psi = -52^\circ$. These comparisons are summarized as follows:

$$\begin{aligned} \text{Tr}^{(1)} \Big|_{\text{computed}} - \text{Tr}^{(1)} \Big|_{\text{measured}} &= -1 \pm 5 \cdot 10^{-6}/^\circ\text{C} \\ \text{Tr}^{(2)} \Big|_{\text{computed}} - \text{Tr}^{(2)} \Big|_{\text{measured}} &= -3 \pm 5 \cdot 10^{-9}/^\circ\text{C}^2 \end{aligned}$$

The main reason for these discrepancies is thought to be the assumption that $s'_{56} = 0$, which yields a slightly inaccurate frequency equation. Considering the various relations involved in Love's analysis, it is easy to show that a first-order frequency correction should be proportional to s'_{56} . But the dependence of the proportionality constant upon the angle of cut and upon the dimensional ratio is unknown, so that its influence on the temperature behavior cannot be determined.

Piezoelectric coupling

The highest possible piezoelectric coupling is obtained by adding the contributions of the electric fields E_2 and E_3 . The piezoelectric constants involved are d'_{25} and d'_{36} , which have the same sign except within two narrow angular domains. The best electrode pattern is shown in fig. 11, with the restriction that the extension of the vertical metallization over the edges is not desirable, but may be called for by practical considerations.

The various energy densities over the cross section are given by the following relations:

$$\text{mechanical} : u_m = \frac{1}{8ab} C \tau^2$$

$$\text{electrical} : u_e = \frac{1}{8ab} \epsilon c_0 \Psi_0^2$$

$$\text{piezoelectric} : u_p = \frac{1}{8ab} \iint_{\text{cross section}} \left[d'_{25} T_5 \frac{\partial \Psi}{\partial y} - d'_{36} T_6 \frac{\partial \Psi}{\partial z} \right] dy dz$$

where $0 \leq \Psi \leq \Psi_0$: electrostatic potential

c_0 : normalized static capacitance per unit length

Combining the various equations derived previously leads to the relation giving the static piezoelectric coupling factor:

$$k^2 \equiv \frac{u_p^2}{u_m u_e} = \frac{12}{\sum_0} \frac{a}{b} \frac{s'_{55} \left[\frac{d'_{25}}{s'_{55}} J_1 + \frac{d'_{36}}{s'_{66}} J_2 \right]^2}{\epsilon c_0}$$

where

$$J_1 = \sum_3 + \frac{1}{a} \int_{a'}^{a''} (1 - \sum_2) \Psi \Big|_{y=b} dz - \frac{1}{4ab} \iint_{\text{cross section}} (1 - \sum_1) \Psi dy dz$$

$$J_2 = \frac{1}{4ab} \iint_{\text{cross section}} \sum_1 \Psi dy dz$$

and

$$\begin{aligned} \sum_0 &= 1 - \frac{192}{\pi^5} \sum_{n=0}^{\infty} \frac{\tanh \frac{(2n+1)\pi a}{2}}{(2n+1)^5} \\ \sum_1 &= \frac{4}{\pi} \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} \frac{\cosh \frac{(2n+1)\pi a}{2}}{\cosh \frac{(2n+1)\pi a}{2}} \cos \frac{(2n+1)\pi \beta}{2} \\ \sum_2 &= \frac{8}{\pi^2} \sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} \frac{\cosh \frac{(2n+1)\pi a}{2}}{\cosh \frac{(2n+1)\pi a}{2}} \\ \sum_3 &= \frac{a'}{a} - \frac{16}{\pi^3} \sum_{n=0}^{\infty} \frac{1}{(2n+1)^3} \frac{\sinh \frac{(2n+1)\pi a}{2}}{\cosh \frac{(2n+1)\pi a}{2}} \frac{a'}{a} \end{aligned}$$

The electrostatic potential Ψ , and the normalized static capacitance c_0 , are determined by applying a method based on the use of conformal trans-

formations, to obtain a uniform electrostatic field. It is thereby assumed that the crystal is electrically isotropic, and the influence of the external electrostatic field is neglected. This method, which makes use of Jacobi elliptic functions and of elliptic integrals of the first kind, has been described in an earlier paper devoted to flexural resonators⁷.

The determination of the dynamic coupling factor calls for integration over the length of the bar, which is assumed to be only partially metallized. The edge effects along the length are neglected, and the stress distribution is taken as sinusoidal.

The static and motional capacitances, respectively C_0 and C_1 , as well as the dynamic coupling factor k_d , are given by the following relations:

$$\begin{cases} C_0 = \epsilon l c_0 \\ C_1 = \mu l c_1 \end{cases} \quad \frac{C_1}{C_0} \approx k_d^2 = \mu k^2$$

where

l : metallized length (symmetrical with respect to the center of the bar)

$c_1 = \epsilon k^2 c_0$: normalized motional capacitance

$$\mu = \frac{4}{\pi} \frac{\sin^2(\frac{\pi l}{2L})}{(\frac{\pi l}{2L})}$$

The complexity of the various transformations involved precludes any determination of the static coupling factor by analytical integration. A computer program has been used for determining the electrostatic potential on a regular array and for performing numerical integrations.

Curves have been computed for the case $\varphi = -53^\circ$, with the vertical metallization limited to the edges ($a'' = a$). Fig. 12 gives the normalized static capacitance c_0 , fig. 13 the normalized motional capacitance c_1 , and fig. 14 the square of the static piezoelectric coupling factor. These curves provide the information needed to determine the static and motional capacitances of DT-cut torsional resonators.

Torsional tuning-fork

A certain analogy between flexure and torsion suggests that two torsional bars can be coupled to form a torsional tuning-fork. This view is supported by experimental evidence, showing that such resonators essentially retain the basic properties of torsional bars.

The analysis of torsional tuning-forks may prove to be even more difficult than that of their flexural counterparts. This is due to the predominantly out-of-plane particle motion and to the more complex stress distribution, and no attempt has been made to achieve this goal. Instead, an empirical frequency relation has been derived, based on the definition of an "equivalent clamping-plane", as represented in fig. 15.

It is found experimentally that the tuning-fork torsional frequency is nearly equal to that of a torsional bar whose half-length is defined in that figure. The frequencies computed with this empirical relation have been compared to the measured frequencies of several torsional tuning-forks, having quite different angles of cut and dimensional ratios. The corresponding frequencies are found to differ by less than 6%.

It is interesting to note that the same equivalent clamping-plane may be defined for flexural tuning-forks, which are then to be considered as two clamped bars of length $L/2$. In this case, computed and measured frequencies agree within 4%.

The temperature behavior of DT-cut torsional tuning-forks has been measured, and found to differ slightly from that of DT-cut torsional bars. For instance, the second-order temperature coefficient of optimized tuning-forks is about $-20 \cdot 10^{-9}/^\circ\text{C}^2$, compared to $-15 \cdot 10^{-9}/^\circ\text{C}^2$ for torsional bars.

Tuning-forks have several advantages over torsional bars. They are suitable for batch-processing, easier to mount, and exhibit a better shock resistance. Furthermore, the electrode pattern may be chosen so as to build a three-terminal device, thus simplifying the design of stable low-power oscillators.

It is felt that torsional tuning-forks may be considered as an alternative to flexural tuning-forks, in particular for wrist-watch applications.

Conclusions

This paper gives the equations necessary for computing the properties of torsional quartz resonators whose length is oriented along the X-axis, and having a rectangular cross section. It shows that DT-cut torsional resonators exhibit a temperature behavior similar to the one of DT-cut shear resonators, and that they can be built as torsional tuning-forks.

The analysis presented is not restricted to quartz crystals, but may be applied to similar materials, such as berlinite, or to materials of higher crystal symmetry.

Acknowledgement

The author wishes to express his thanks to C. Bourgeois, who wrote the computer programs needed for determining the piezoelectric coupling factor.

References

1. E. Giebe and E. Blechschmidt, "Ueber Drillungsschwingungen von Quarzstäben und ihre Benutzung für Frequenznormale", *Hochfrequenztechnik und Elektroakustik*, Band 56, Heft 3, September 1940, pp. 65-87.
2. I.G. Vasin and P.G. Pozdnjakov, "Quartz Resonator", US Patent 3 376 439, awarded April 2, 1968.

3. P.G. Pozdnjakov, "Quartz Crystals Resonating in Torsion", Soviet Physics Crystallography, Vol.15, No1, July-August 1970, pp. 63-87.
4. A.E.H. Love, "A Treatise on the Theory of Elasticity", Dover Publications (S 174), 1944, pp. 310-328.
5. L. Landau and E. Lifchitz, "Théorie de l'Elasticité", Editions MIR, Moscou 1967.
6. R. Bechmann, A.D. Ballato and T.J. Lukaszek, "Higher-Order Temperature Coefficients of the Elastic Stiffnesses and Compliances of Alpha-Quartz", Proceedings of the IRE, August 1962, pp. 1812-1821.
7. J. Hermann, "Determination of the Electro-mechanical Coupling Factor of Quartz Bars Vibrating in Flexure or Length-Extension", Proceedings 29th Annual Symposium on Frequency Control, May 1975, pp. 26-34.

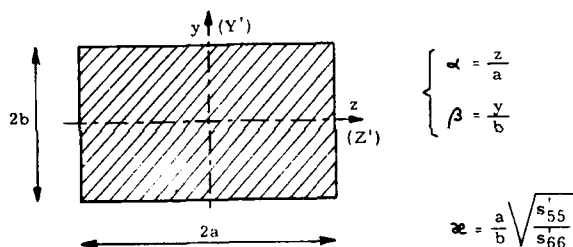


Fig. 1 Cross section of a torsional bar

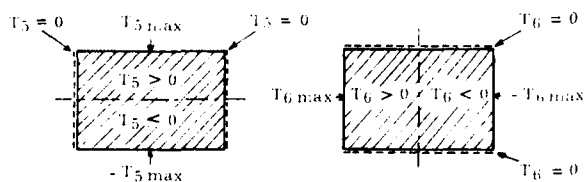


Fig. 2 Stress symmetries

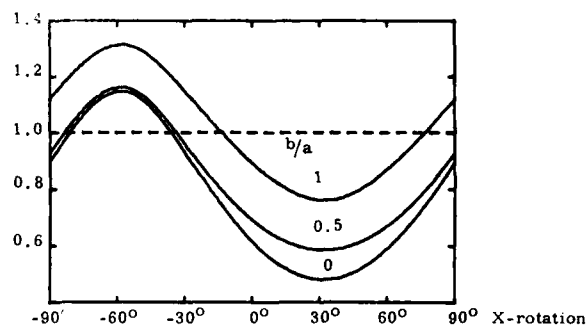


Fig. 3 Ratio $T_6 \max/T_5 \max$

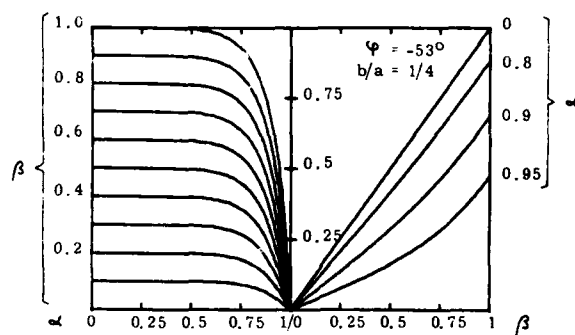


Fig. 4 Normalized stress $T_5/T_5 \max$

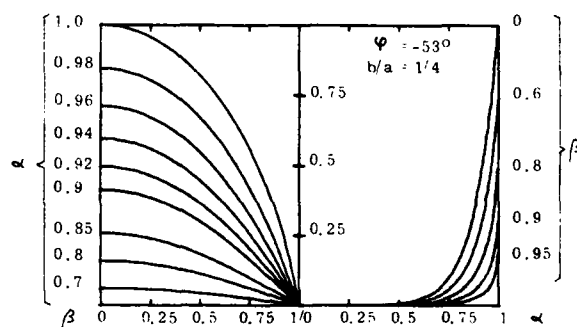


Fig. 5 Normalized stress $T_6/T_6 \max$

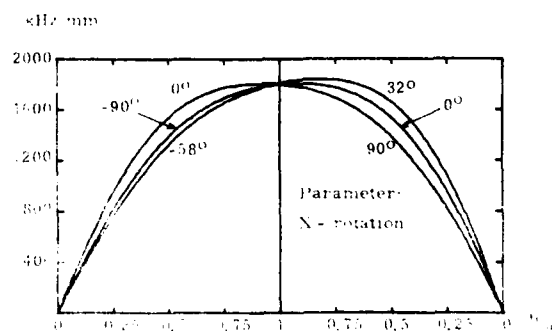


Fig. 6 Frequency constant

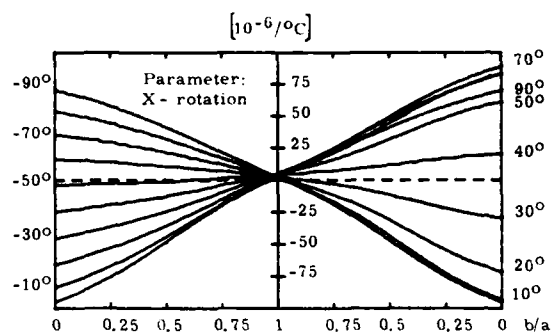


Fig. 7 First-order temperature coefficient $T_f^{(1)}$

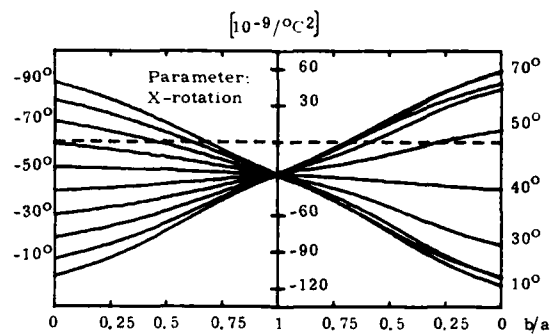


Fig. 8 Second-order temperature coefficient $T_f^{(2)}$

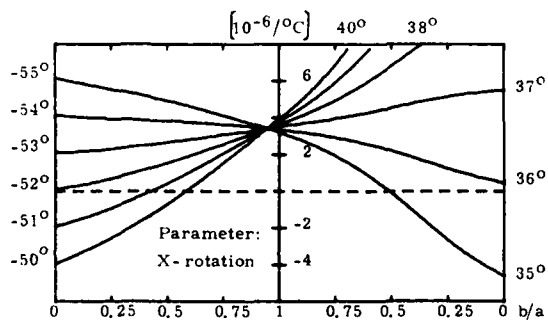


Fig. 9 DT/CT - cuts: $T_f^{(1)}$

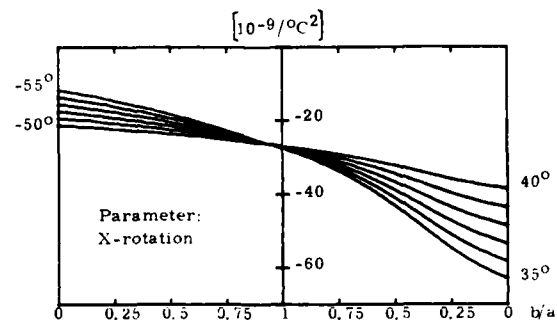


Fig. 10 DT/CT - cuts: $T_f^{(2)}$

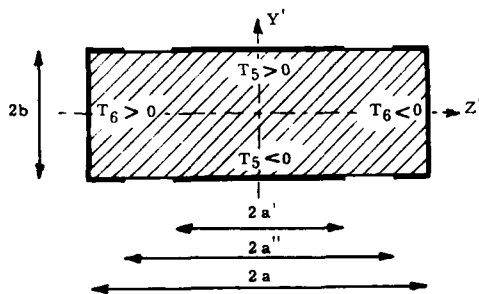


Fig. 11 Metallization pattern

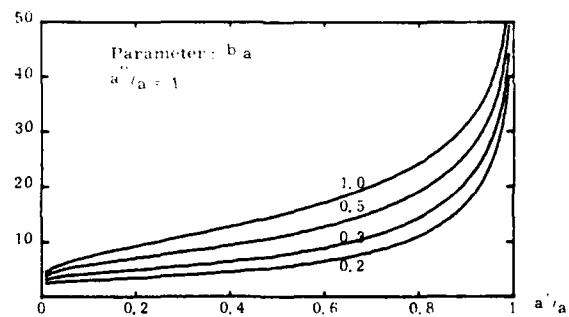


Fig. 12 Normalized static capacitance c_0

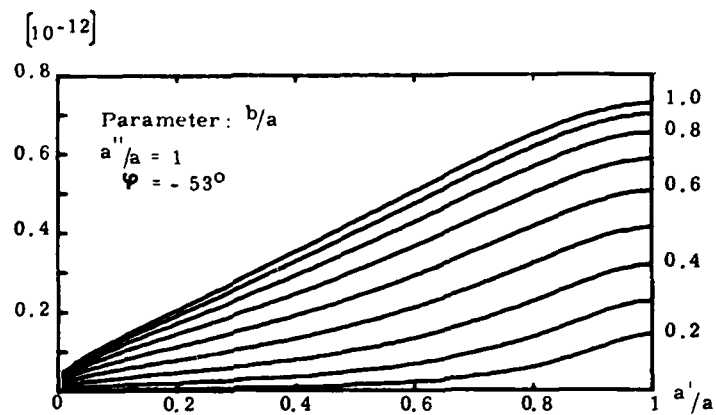


Fig. 13 Normalized motional capacitance c_1

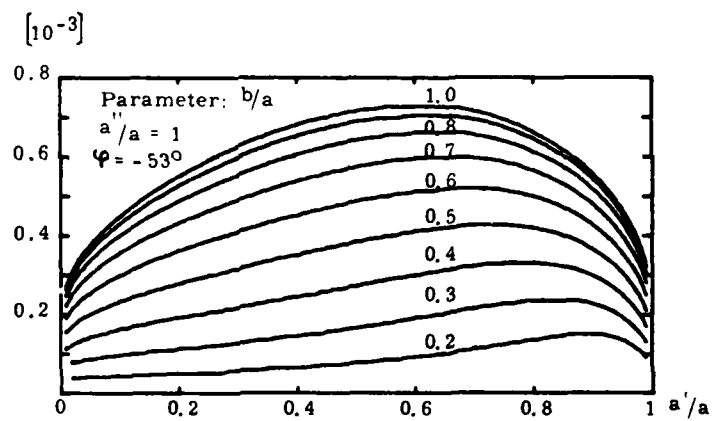


Fig. 14 Static piezoelectric coupling factor k^2

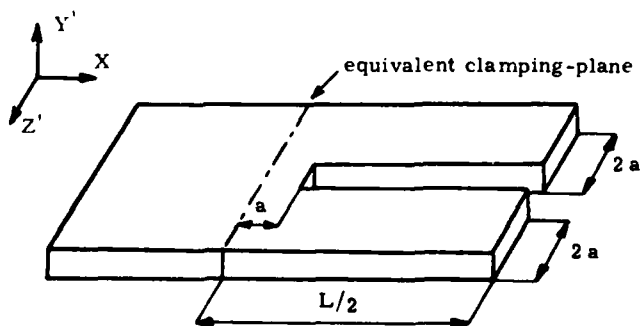


Fig. 15 Torsional tuning-fork

QUARTZ CRYSTAL ACCELEROMETER INSENSITIVE TO TEMPERATURE VARIATION

Morio Onoe
Institute of Industrial Science
University of Tokyo
Roppongi, Tokyo, Japan

Kaoru Furusawa
Meisei Electric Co., Ltd.
Bunkyo-ku, Tokyo, Japan

Sumiyuki Ishigami
Takeshi Sase
and
Mitsuru Sato
Tokyo Denpa Co., Ltd.
Ota-ku, Tokyo, Japan

Summary

An accelerometer has been developed, which utilizes force sensitivity of quartz crystal resonators. Two identical rectangular resonators are stacked together with a small gap and clamped along one edge as a cantilever. A mass is attached to the opposite edge. Each resonator is driven by an oscillator circuit and oscillates at its own resonant frequency of thickness-shear mode. The output of both oscillators is mixed to yield the beat frequency, which is proportional to the acceleration and hence the desired output of the accelerometer. Since the output is the frequency, the digital data processing can be easily applied. For telemetry, a FM transmitter of around 170 MHz is modulated by the beat frequency output. The size of the whole package is 40×20×20 mm.

The accelerometer shows an uniaxial force sensitivity. This is because only the force component perpendicular to the major surfaces of resonators stretches one resonator and compresses another resonator. This causes opposite shifts in oscillator frequencies and yields the beat output. On the other hand, other force components cause the same shift in oscillator frequencies and hence no beat output.

Similar cancellations are obtained for the DC offset due to temperature variation and aging. The temperature variation of the force sensitivity of two resonators, however, cannot be cancelled out. Hence it is required to use resonators, of which force sensitivity is insensitive to temperature variation. Hence in this study a search was done for crystal cuts with zero temperature dependence of force sensitivity, several useful cuts with zero temperature dependence of force sensitivity have been found, which includes BT- and RT-cuts, besides of AT-cut already reported by Dauwater (1972). In the case of a rectangular plates, the temperature dependence is a function of the location of point of application of force. Hence another possibility exists to obtain the zero temperature dependence.

Introduction

Resonance frequencies of quartz crystal plates change by the application of external force. This effect is troublesome in the application of high-precision quartz resonators. On the other hand, the use of this effect can be made to realize a force to frequency sensor. In practical uses of such a sensor, it is important to reduce temperature dependence of the force to frequency sensitivity as well as the frequency itself.

Much work has been done on the effect of external

force on the frequency of a quartz resonator. Ballato and Bechmann studied the sensitivity as a function of the angle of cut and the azimuth angle.^{2,3} Then followed extensive work by Mingins et al.⁴ Ratajski summarized data of rotated Y-cut plates in a chart form.⁵

Theoretical calculation of the sensitivity was done by Keyes and Blair, who assumes isotropy in stress analysis of a crystal plate.⁶ More general treatment was presented by Lee et al.⁷ These calculations using the third order elastic constants determined by Thurston et al.⁸ showed a fair agreement with experimental data.

Only a few data is available for the temperature dependence of the sensitivity. Hammond et al. reported that a large temperature coefficient of the sensitivity of a cylindrical pressure sensor using an AT-cut was substantially reduced by the use of a doubly rotated cut.⁹ Dauwater measured temperature dependence of the force sensitivity of an AT-cut as a function of azimuth angle and found a zero temperature dependence at the angle of 37 degree from the X-axis.¹⁰

In this paper, first the temperature dependence of the sensitivity of other rotated Y-cuts as well as doubly rotated cuts are presented. Several useful cuts and azimuth angles with zero temperature dependence of the force sensitivity are found, which include BT-cut and RT-cut.

Second the temperature dependence of the force sensitivity of a rectangular quartz crystal plate as a function of the location of point of application of force is presented.

Third the development of an accelerometer insensitive to temperature variation and its characteristics are described.

Force to Frequency Sensitivity and Its Temperature Dependence

The normalized coefficient of the force to frequency sensitivity is defined by Ratajski⁵ as follows:

$$K_f = \frac{\Delta f}{F} \frac{nD}{f^2 \eta} \quad (1)$$

where

Δf : change in frequency due to force
F: applied force
n: order of overtone
D: diameter of quartz crystal plate
f: resonant frequency
 η : a parameter which is assumed as unity in this study.

Rotated Y-cuts

Theoretical analysis shows that the normalized coefficient of a rotated-Y cut depends on the azimuth angle, ψ , of applied force in the following manner⁷:

$$K_f(\psi) = P + Q \cos(2\psi) \quad (2)$$

where P and Q are parameters which depend on the angle of cut. Hence the temperature dependence of K_f is fully described by the temperature dependence of two parameters P and Q or, equivalently, $K_f(0)$ and $K_f(90)$.

In the experiment, 10.7 MHz rotated Y-cut circular resonators are used. The diameter is 8 mm. Gold electrode of 2.6 mm in diameter is evaporated in the center. A compressional force of 200 ~500 gram is applied in parallel to the major surface. The temperature dependence is measured in the range from zero to 80°C.

Fig.1 shows a comparison of theoretical and experimental values of K_f for AT and BT cuts as a function of the azimuth angle. A fair agreement is obtained as described in literatures. Fig.2 and 3 show the linear temperature coefficient of $K_f(0)$ and $K_f(90)$ at 35°C, respectively, as a function of the angle of cut. The coefficient of $K_f(0)$ becomes zero at the angle of cut of -55, -43 and +20 degree. On the other hand, the coefficient of $K_f(90)$ becomes zero at the angle of cut of -10, +3, and +20 degree.

The temperature dependence of K_f for any combination of the azimuth angle and the angle of cut can be calculated from data shown in Fig.2 and 3. In practice, the temperature dependence of resonant frequency itself is also important. Hence AT and BT cuts are to be used. Fig.4 and 5 show the temperature dependence of K_f of AT and BT cuts for several values of the azimuth angle. In the case of AT-cut, the azimuth angle around 40 degree yields good temperature characteristics as reported by Dawwater.¹⁰ In the case of BT-cut, a zero azimuth angle yields a good result.

Doubly rotated cuts

A search for cuts with good temperature characteristics of the force sensitivity is extended to doubly rotated cuts. Results are summarized in Table 1. The figure of merit, M, is the ratio of the maximum variation of K_f to the mean value of K_f over the measured temperature range. The values for AT and BT cuts are also included for comparison.

It can be seen RT cut ($\phi=14^\circ 30'$ and $\theta=-33^\circ 48'$), which is one of doubly rotated cuts having zero temperature coefficient of frequency, yields a good result at the azimuth angle of 120 degree. Fig.6 shows theoretical and experimental values of K_f of RT-cut as a function of the azimuth angle. The curve does not exhibit a symmetry around the zero azimuth angle. Fig.7 shows the temperature dependence of $K_f(35)$ and $K_f(120)$. Little improvement in temperature characteristics of force sensitivity is obtained by using doubly rotated cuts.

Rectangular plates

So far the circular plates are used in the experiment. Similar experiments are conducted for rectangular AT cuts in order to see the effect of the shift in location of point of application of force. The size of the plate is 10x15x0.15 mm. Fig. 8 shows three types of rectangular plates and the point of application of force. Fig.9 shows K_f vs the shift in location of point of application of force. The

temperature dependence of K_f at points specified in Fig.9 is shown in Fig.10.

It can be seen in Fig.9 that the force sensitivity becomes zero at certain points. These points may be suitable for mounting a precision crystal resonator. The shift of point of application, however, yield little improvement in temperature characteristics of force sensitivity.

Structure of accelerometer

An accelerometer, of which structure is shown in Fig.11, is developed. Two identical rectangular resonators are stacked together with a small gap and clamped along one edge so that each plate behaves as a cantilever. These resonators are hermetically sealed by bellows. The flexible end of bellows is attached to both the free edges of resonators inside and a weight outside. The inside of bellows is filled with dry nitrogen. Silicon oil fills the space between the bellows and an outer container to increase the damping. Resonators and its mounting material have the same thermal expansion coefficient in order to reduce thermal stress. Each resonator is driven by an oscillator circuit and oscillates at its own resonant frequency of thickness-shear mode. The output of both oscillators is mixed to yield the beat frequency, which is proportional to the acceleration.

The accelerometer exhibits an uniaxial force sensitivity. This is because only the force component perpendicular to the major surfaces of resonators stretches one resonator and compresses another resonator. This causes opposite shifts in oscillator frequencies and yields the beat frequency output. On the other hand, other force components cause the same shift in oscillator frequencies and hence no beat output. Similar cancellations are obtained for the DC offset due to temperature variation and aging.

The temperature variation of the force sensitivity of two resonators, however, cannot be cancelled out. Hence it is important to use resonators of which force sensitivity is insensitive to temperature variation. The previous study suggests the use of either AT-cuts with the azimuth angle of 40 degree or BT-cuts with zero azimuth angle. Although AT-cuts exhibit a little better characteristics, BT-cuts are used in the following, because a matched pair is easily obtained with less accurate control of angles.

A single coaxial cable is connected to the accelerometer. The cable is used for both DC power supply and the output of beat frequency. For telemetry, the beat frequency output modulates a FM transmitter of around 170 MHz as shown in Fig.12. The size of the whole package is 40x20x20 mm. Fig.13 shows a whole view and a cut away view of a typical package.

Characteristics of accelerometer

Typical characteristics of the accelerometer are summarized in Table 2. The maximum frequency can be increased by reducing the weight, which however, reduces the force sensitivity. Fig.14 shows temperature variation of two resonators without applied force and its residual beat output. Fig.15 shows frequency characteristics of the accelerometer. Fig.16 shows uniaxial directivity of the accelerometer. Fig.17 shows the fluctuation in acceleration and deceleration of a motor measured by telemetry.

Conclusion

An accelerometer is developed, which utilizes

force sensitivity of quartz crystal resonators. Two identical resonators combined in a cantilever form, so that the uniaxial directivity is obtained. All the variations except the temperature variation of the force sensitivity can be cancelled out. Hence a search was done for crystal cuts with zero temperature dependence of the force sensitivity. Several useful cuts are found, which include BT-cut and RT-cut, besides of AT-cut previously reported.

Reference

1. V.E.Bottom : Am. Mineral. 32, 590-591 (1947)
2. A.D.Ballato and R.Bechmann : Proc. IRE 48, 261-262 (1960)
3. A.D.Ballato : Proc. Freq. Cont. Symp., 89-114 (1960)
4. C.R.Mingins : Proc. Freq. Cont. Symp., 46-76 (1962), 51-87 (1963), 50-69 (1966)
5. J.M.Ratajski : IBM J. Res. and Dev. 12, 92-99 (1968)
6. R.W.Keyes and F.W.Blair : Proc. IEEE 55, 565-566 (1967)
7. P.C.Y.Lee, Y.S.Wang and X.Markenscoff : J. Acoust. Soc. Am. 57, 96-105 (1975)
8. R.N.Thurston, H.J.McSkimin and P.Andreatch : J. Appl. Phys. 37, 267-275 (1966)
9. D.L.Hammond and A.Benjaminson : IEEE Spectrum 53-58 (April, 1969)
10. C.R.Dawwalter : Proc. Freq. Cont. Symp. 108-112 (1972)

Appendix

Azimuth Angles which yield zero force sensitivity

In the design of a high precision quartz resonator, it is interesting to know azimuth angles yielding zero force sensitivity. This is because the mounting at these angles may reduce mechanical disturbances caused by a holder.

Fig.1 shows the azimuth angles of about 60 and 120 degree can be used for two and four points mounting of an AT-cut plate.

The azimuth angles yielding zero force sensitivity for three points mounting was experimentally studied by Sasaki et al.^{A1} Fig.A1 shows the geometry of three points mounting. They found the angles of zero force sensitivity at both ($\alpha=75^\circ$, $\psi=0^\circ$) and ($\alpha=36^\circ$, $\psi=90^\circ$).

These angles can be calculated from the data shown in Fig.2 and 3 and the following equation:

$$K_f = P(1 + \frac{1}{2}\tan\alpha) + Q\cos(2\psi)(1 - \frac{1}{2}\tan\alpha)$$

Fig.A2 shows calculated K_f of an AT-cut as a function of azimuth angle. The parameter is α . It is seen zero force sensitivity can be obtained in the range of α from 0 to 39 degree, if ψ is appropriately chosen. Another combination ($\alpha=78^\circ 35'$, $\psi=0^\circ$) also yields zero sensitivity.

Temperature dependence of K_f for various combinations of angles yielding zero sensitivity at 30°C is given in Fig.A3. It is seen that the two points mounting, ($\alpha=0^\circ$, $\psi=57^\circ 34'$) exhibits a minimum sensitivity over a wide temperature range.

Reference

- A1 E. Sasaki, H. Kojima and S. Hattori: Nat. Conv. Inst. Electronics and Elec. Comm. Eng. Japan, No.95 (1975)

Table 1 Comparison of figure of merit of doubly rotated cuts and AT and BT cuts.

ϕ	θ	ψ	F.M.
$7^\circ 03'$	$-48^\circ 18'$	70°	0.074
		155°	0.131
$11^\circ 43'$	$-32^\circ 58'$	50°	0.076
		145°	0.101
$14^\circ 30'$	$-33^\circ 48'$	35°	0.099
		120°	0.045 ←
$12^\circ 20'$	$28^\circ 12'$	95°	0.113
		170°	0.060
$13^\circ 51'$	$39^\circ 48'$	150°	0.192
AT		0°	0.095
		90°	0.273
		40°	0.019
BT		0°	0.043 ←

Table 2 Typical characteristics of accelerometer

LINEARITY	0.2 % of full scale
HYSTERESIS	0.2 % of full scale
MAXIMUM FREQUENCY	100 Hz for full scale of 10 G
SENSITIVITY	500 Hz/G for full scale of 10 G
RESOLUTION	2×10^{-3}
TEMPERATURE COEFFICIENT OF SENSITIVITY	-0.01 %/°C

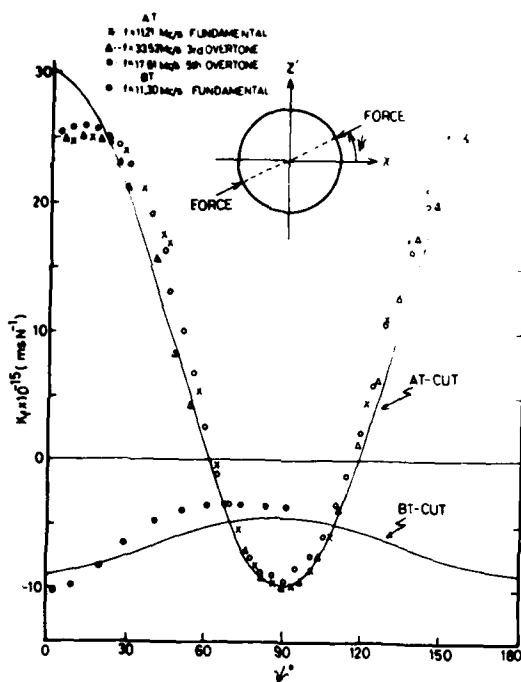


Fig. 1. Normalized coefficient of the force sensitivity of AT and BT cuts as a function of the azimuth angle of applied force

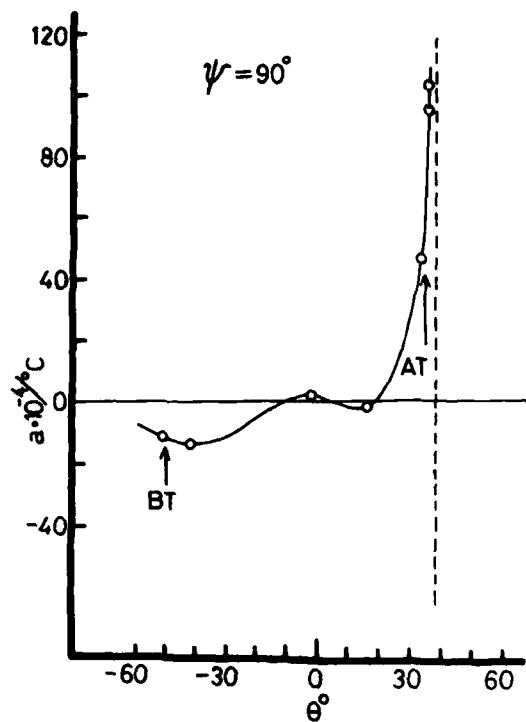


Fig. 3. Linear temperature coefficient of $K_f(90)$ as a function of the angle of cut

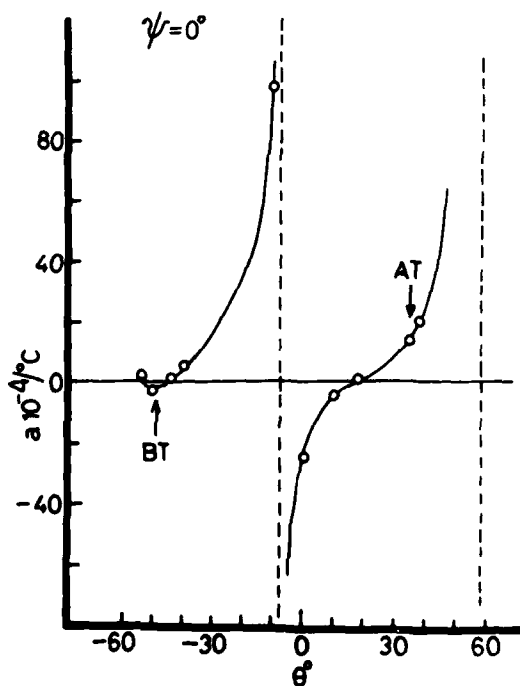


Fig. 2. Linear temperature coefficient of $K_f(0)$ as a function of the angle of cut.

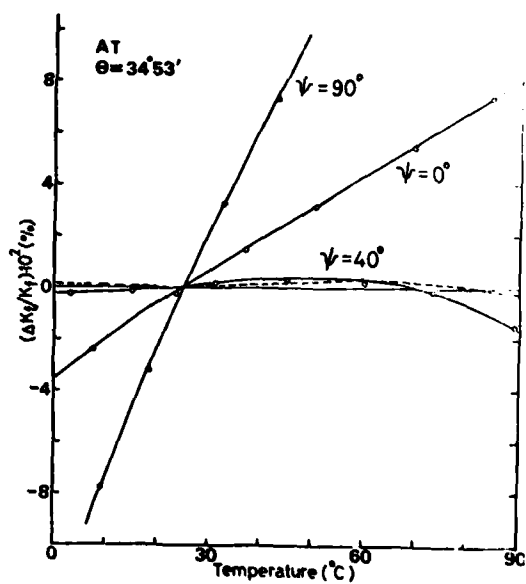


Fig. 4. K_f vs. temperature for AT cut

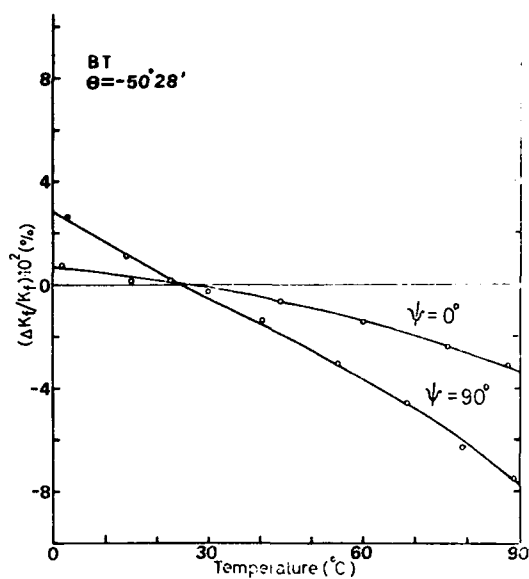


Fig. 5. K_f vs. temperature for BT cut

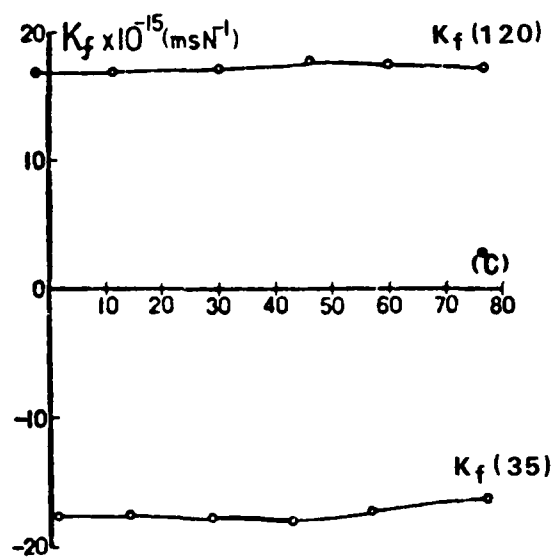


Fig. 7. Temperature dependence of $K_f(35)$ and $K_f(120)$ of RT-cut

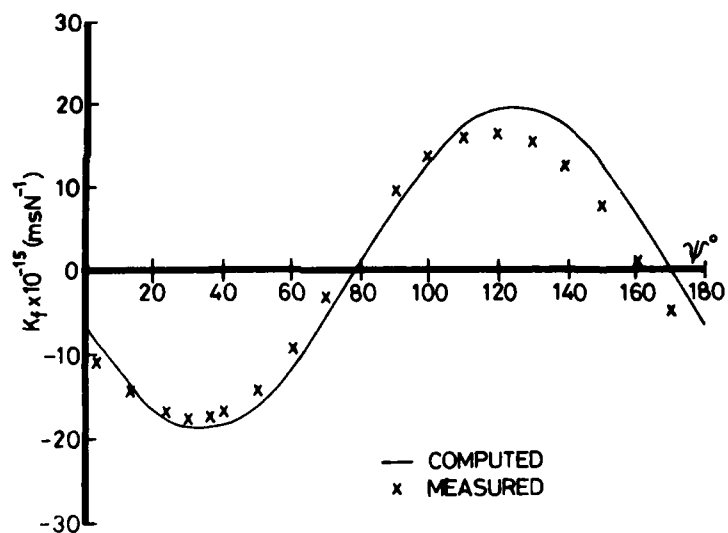


Fig. 6. K_f vs. azimuth angle for RT cut

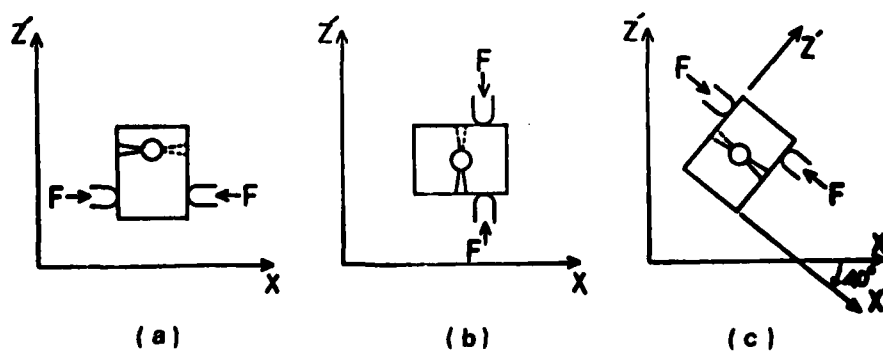


Fig. 8. Application of force to rectangular AT-cuts

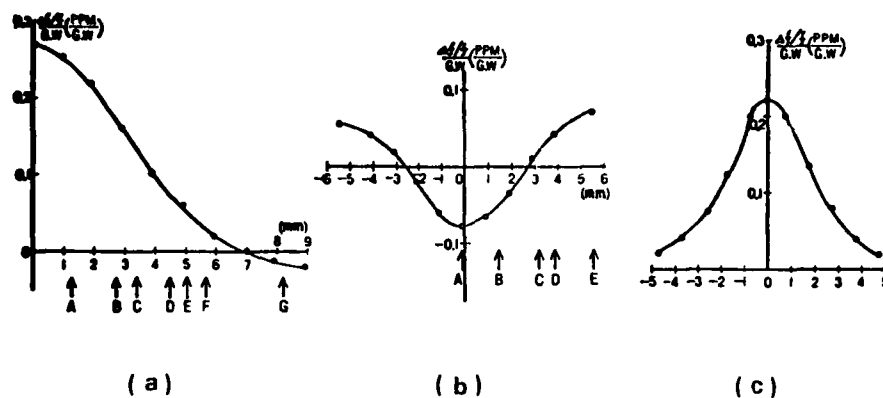


Fig. 9. K_f vs. shift in location of point of application of force

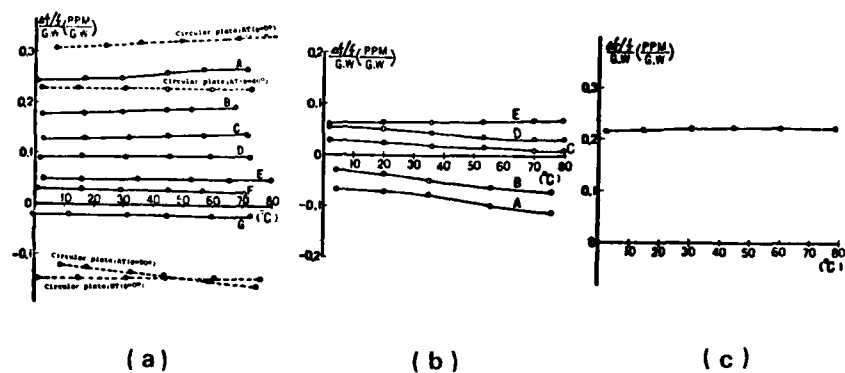


Fig. 10. K_f vs. temperature at various points specified in Fig. 9

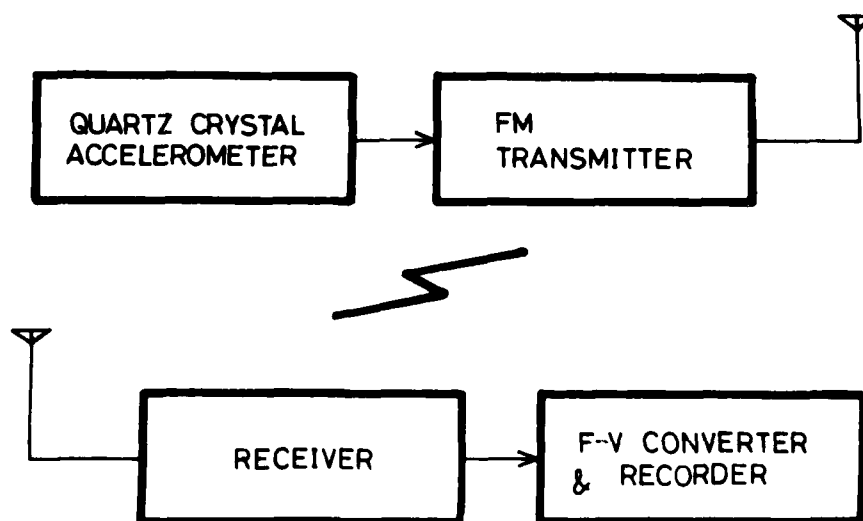


Fig. 12. Arrangement for telemetry

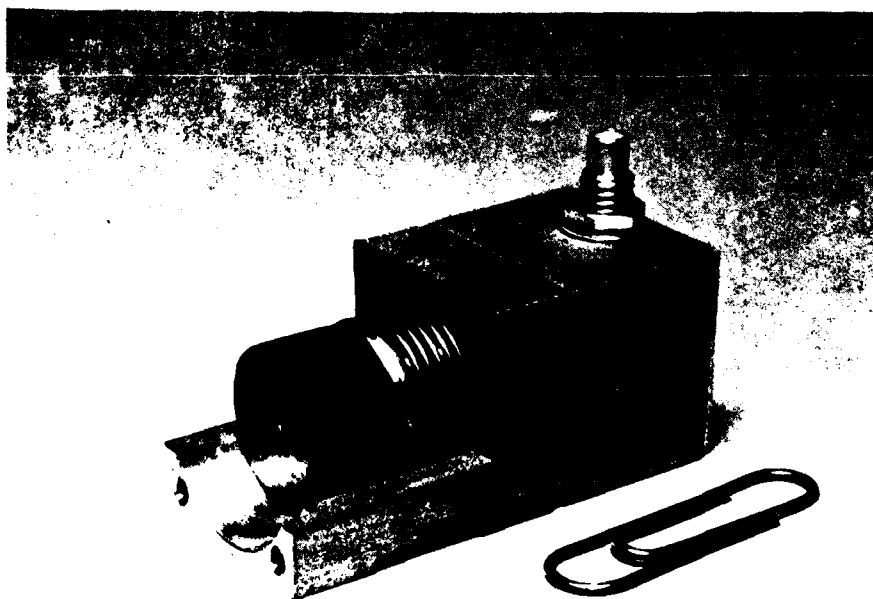
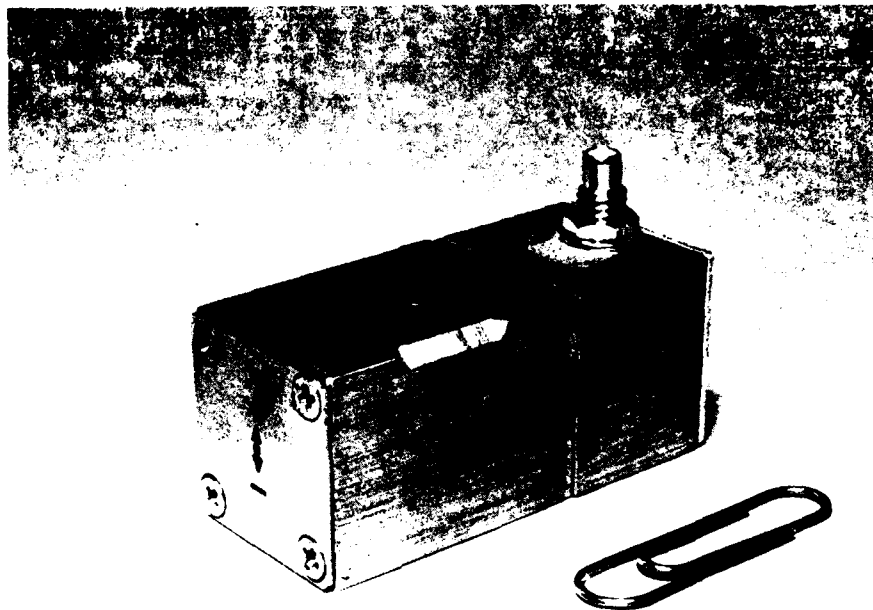


Fig. 13. Whole and cut away views of accelerometer

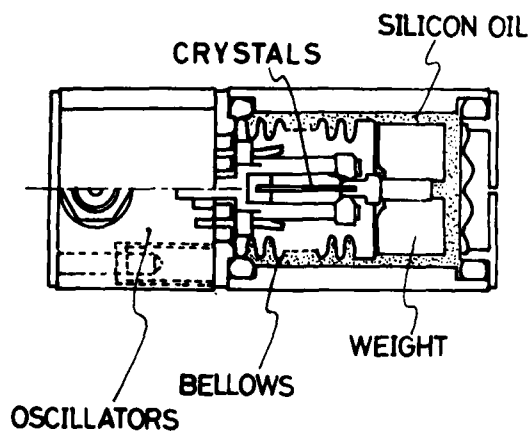


Fig. 11. Schematic view of accelerometer

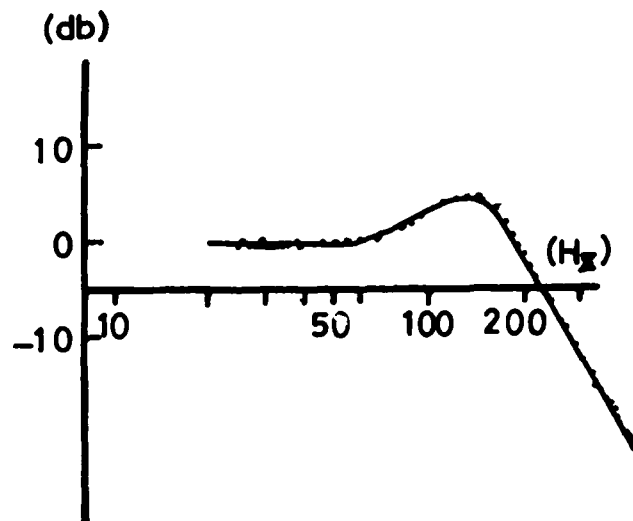


Fig. 15. Frequency characteristics of accelerometer

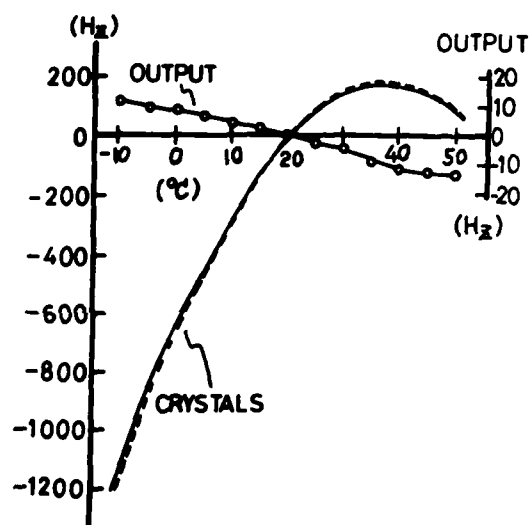


Fig. 14. Temperature variation of oscillating frequencies of two resonators and its residual beat frequencies output

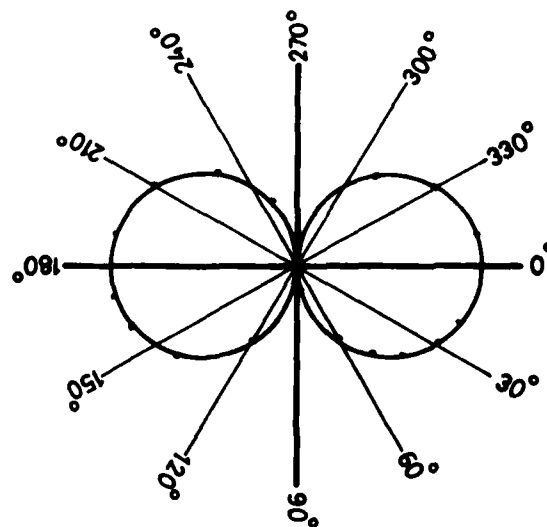


Fig. 16. Directivity of accelerometer

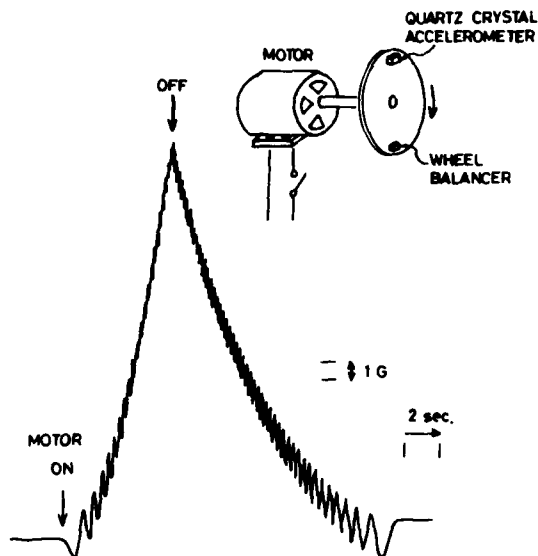


Fig. 17. Fluctuation in acceleration and deceleration of a motor measured by telemetry

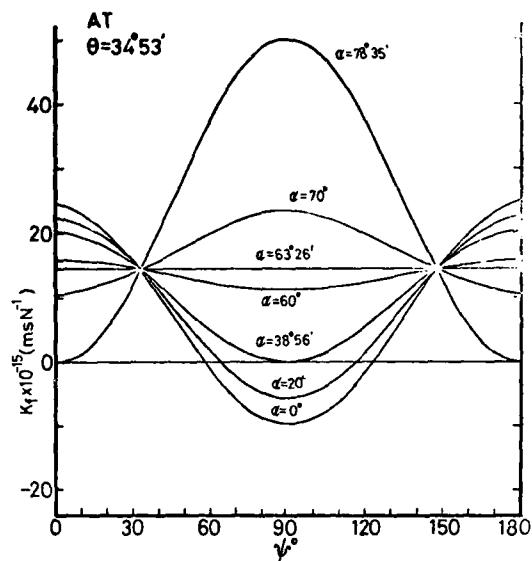


Fig. A2. K_f vs. azimuth angle ψ of AT-cut (the parameter is the angle α)

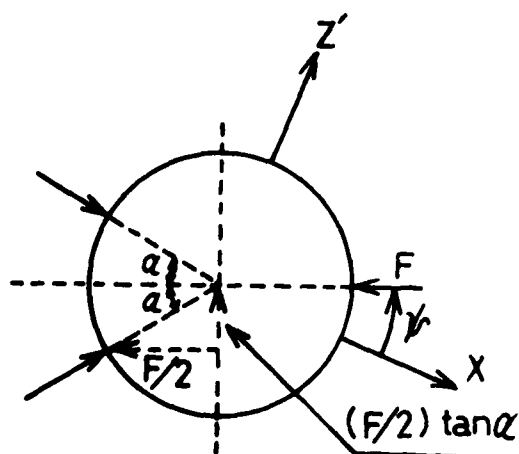


Fig. A1. Geometry of three points mounting

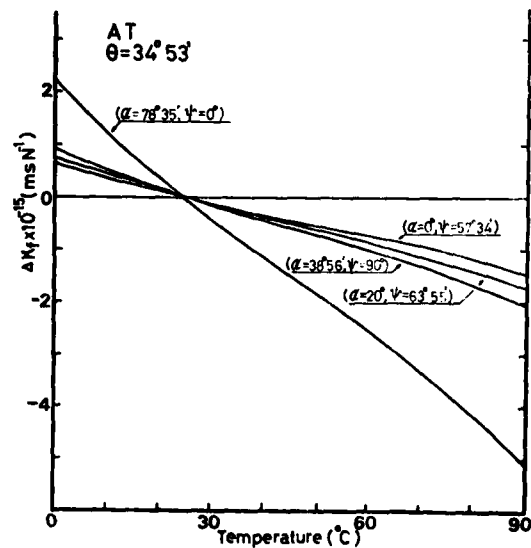


Fig. A3. Temperature dependence of K_f for combinations of angles yielding zero force sensitivity at 30°C

USING THE X-Y FLEXURE WATCH CRYSTAL

AS A PRESSURE-FORCE TRANSDUCER

Alan Genis - D.E. Newell
Northern Illinois University
DeKalb, Illinois

Summary

This paper describes a method of utilizing a 32.768KHz X-Y flexure watch crystal as a pressure-force transducer. It was determined that the frequency of oscillation decreases approximately - 0.4Hz/psi, (coefficient varies with crystal design), and the slope of the frequency versus pressure curve is approximately constant, beyond an initial threshold.

The X-Y flexure resonator was enclosed in a compartment with one side being formed by a diaphragm composed of a synthetic rubber impregnated fabric. This compartment was filled with nitrogen gas at atmospheric pressure. A piston was positioned on the external side of the diaphragm and as a force was applied to the piston, the volume of the crystal enclosure would decrease. This decrease in volume would increase the nitrogen pressure in the crystal compartment increasing the viscous damping of the resonator and thus decreasing the crystal frequency.

The crystal was placed in an oscillator circuit, generating a frequency that was a function of the force applied to the piston. Electronic circuitry was generated that would convert this frequency shift to a digital indicator, indicating the force in lbs. Self compensating circuitry was developed to eliminate any temperature or aging effects which would cause a frequency shift other than the shift caused by viscous damping. Two models of the transducer have been generated.

The design considerations and test results obtained from the two models will be discussed.

Introduction

Various reports have been published and devices generated in the past fifty years that utilize a piezoelectric material for the detection of transient forces or pressure measurement. The original suggestion that pressure could be recorded as a function of time using piezoelectric crystals has been credited to J.J. Thompson.

Most of the literature describes devices that utilize the piezoelectric material in the static sense, that is, the crystal experiences some form of mechanical pressure or deformation resulting in a charge displacement and induced voltage that is amplified and processed to give the magnitude of the literature published about static transducers, there are relatively few articles that describe the use of piezoelectric crystals in an oscillatory sense for the use as force transducers. Of this small

number of articles, most describe devices that experience a physical deformation of the crystal caused by differential air loading or mechanical forces applied to the crystal to force the crystal to deviate from its resonant frequency.

Brown and Harris¹ described a method of determining the pressure coefficient of a BT cut quartz crystal, and noticed that this frequency shift was linear with changing pressure with a coefficient of .408ppm/lbs/in².

In 1951 Roberts and Goldsmith² described two methods of pressure measurement utilizing an oscillatory type of quartz crystal. The first method explains the effects of air loading on the Q of the crystal. It is noticed in this study that for a crystal vibrating in extension or flexure, the Q change is inversely proportional to air pressure and for elements operating in the shear mode the change in Q is inversely proportional to the square root of the pressure.

In this experiment the crystal experienced pulses of fixed amplitude under conditions of varying environmental pressures and the ratio of the decay times used as the means of determining the pressure. The second approach described a method using differential air loading on the quartz crystal. With thickness to length ratios of 80 for AT cut quartz crystals, frequency changes of 1400 ppm/atmosphere were achieved. For an oscillator which has a stability of 1ppm, an accuracy of .7 millibar can be achieved for a pressure shift of one atmosphere.

Pacey³ generated a piezoelectric manometer using a 200KHz DT cut quartz crystal operating in the face shear mode. The crystal was placed in a Pierce oscillator and measurement of the anode current was seen to be proportional to the pressure.

It is the purpose of this report to describe the effect of viscous damping on the resonant frequency exhibited by the X-Y flexure watch crystals and to utilize this effect as a force transducer.

Experimental Procedure

The crystal used is a Type TX-3 X-Y flexure manufactured by the C.T.S. Knights Company. In order to allow the crystal to experience varying environmental pressures, a small portion of the crystal can was removed. The crystal was placed in a pressure chamber and connected to an oscillator, that was external to the chamber, via vacuum feed-throughs. The pressure was varied within the chamber from one atmosphere to 50 p.s.i., while a period count was taken on a Model 5245L Hewlett Packard

frequency counter. A thermocouple was placed within the chamber to monitor temperature to assure that the readings were taken at a constant temperature to eliminate unwanted temperature effects. The data obtained is tabulated in Table 1 and a plot of frequency vs. pressure is shown in Fig. A.

From the data obtained, it is seen that for the X-Y flexure experiencing the change in pressure, the frequency change has an average value of .424 Hz per P.S.I. and this appears to be linear as seen from the graph of Fig. A. It was because of this linearity and the magnitude of the coefficient that suggested a device could be designed to utilize the effect.

The first problem was to develop a transducer that would translate a physical force to a varying environmental pressure on the crystal. The first transducer developed is shown in Fig. B. It consisted of an aluminum housing with a brass piston acting on a rubber diaphragm. A teflon sleeve was inserted between the aluminum and brass to reduce the friction. The second transducer developed is shown in Fig. C. This transducer comprised a Beryllium Copper convolute of a vacuum bellows sealed to copper clad G10 board. One side of the transducer was notched and the crystal epoxied in place. The transducer was then backfilled with dry nitrogen to a pressure of one atmosphere. Other transducers were generated using the bellows arrangements. Some consisted of multiple bellows connected by small tubing to minimize dead air space and maximize working area of the transducer. Another model was generated that consisted of a small air space bounded by a flexible rubber membrane and the bellows was filled with liquids to maximize the working volume of the transducer. See Fig. D.

All of these transducers proved unsatisfactory for the application needed, since they exhibited second order effects possibly due to the spring constants of the bellows and rubber diaphragm. The transducer used in the device was manufactured by Bellofram Corp. This transducer consisted of a rolling sock arrangement on a piston within a housing. See Fig. E. The advantage to the transducer is that it maintains a constant area regardless of the displacement of the piston. It exhibited no restoring forces, except those produced by the comparison of the nitrogen within the transducer. The bellofram requires little force to overcome its own friction.

The quartz crystal was mounted in a holder and mounted in the transducer. The transducer was mounted in a test stand and 50 gram weights were placed on the transducer. Period counts were taken each weight added. Thus the coefficient of the transducer was determined.

Once the coefficient of the transducer had been determined, circuitry had to be developed to utilize the coefficient. The crystal oscillator consisted of a RCA 4007 CMOS inverter biased linearly. The coefficient was multiplied by means of a phase locked loop. Fig. F is a block diagram of the frequency multiplier circuitry, offset oscillator, timing oscillator, gating and readouts.

The nominal frequency of the sensor oscillator is 10,000 Hz. The sensor oscillator will experience a change in resonant frequency due to the viscous damping when a force is applied to the transducer. The frequency of the oscillator will be

$$f_{osc} = f_o - \frac{F}{A} k$$

when f_{osc} = frequency of crystal with a force applied

F = force applied to the transducer

A = cross-sectional area of the transducer

k = pressure coefficient of crystal

f_o = resonant frequency

The f_{osc} is then phase locked to a voltage controlled crystal oscillator with a divider between the VCO and the phase detector, so that when a lock is achieved the VCO will be operating at $128 * f_o$. The VCO frequency is then mixed with the offset oscillator and the difference frequency is generated. Since the coefficient of the transducer is not unity, a gate time had to be generated that would convert the beat frequency of the offset oscillator and the multiplied sensor frequency to a number that would be displayed as the proper magnitude in the units chosen. It was desirable that the gated beat frequency be 10KHz. Then if the most significant bit was disregarded, as the sensor frequency decreased with an increase of pressure on the transducer, the beat frequency would increase and be gated to display the proper magnitude. To determine the proper gate time the following equation was used.

$$t_{on} [t_{off} - m (f_o - \frac{F}{A} k)] = 10,000$$

Once the gating time was known, the offset oscillator frequency can be calculated by

$$[f_{off} - m f_{osc}] t_{on} = 10,000$$

$$f_{off} = \frac{10 \times 10^3}{t_{on}} + m f_{osc}$$

The gating signal was supplied by a crystal oscillator with a frequency of 2.654373 MHz and divided by 2^{20} to generate the needed gating pulse. The device generated is shown in Fig. G. Its operating range was from zero to 200 pounds with an accuracy of .1 pound. This model required external power, a long warm-up period and experienced drift to the fact that the enclosure acted as an oven which caused the crystal frequencies to drift and thus required constant tuning of the offset oscillator. To solve some of the problems associated with the first unit, a second unit was generated. A block diagram is shown in Fig. H.

A new more compact Bellofram was used and the coefficient was determined in the same manner as the first model. The data is shown in Table 2. Because of the drift, some method had to be devised to compensate for the errors. An automatic zero loop was included in the circuit. The loop consisted of a second set of registers which received that data from the mixer. These registers feed the data into comparators which were hard wired to binary 10000. Whenever the difference frequency was not equal to 10000Hz coming out of the gate, the comparators responded with logic levels that corresponded to the frequency being "less than" or "greater than" 10 KHz. These steering outputs were directed thru an interface to the charge pump. The charge pump controlled

the voltage on the varactor in the VCXO which was used as the offset oscillator shown in figure K. Thus when the loop was closed the comparators always maintained the offset oscillator at a frequency that was 10 KHz greater than the frequency of the multiplied sensor oscillator. Since the comparators would always read "greater than" when a force was applied to the sensor or "less than" when the registers were dumped, gating was utilized to open the loop under these conditions, as shown in figure J.

CMOS integrated circuits were used wherever possible to minimize power consumption. The completed unit is shown in figure L. The second model was hand held, capable of operating for eight hours off of its batteries, or off of an external power supply for extended periods. The device has an operating range of 0 to 200 pounds and maintains an accuracy of .1 pound over that range. The accuracy seems limited by the short term stability of the crystals used in the system, and some improvement could be achieved by making the time base oscillator a TCXO, thus minimizing any variation in the gating time. Possible applications for this type of sensor include barometers, depth gages, vacuum gages and a digital bathroom scales.

Table 1

PRESSURE VS. FREQUENCY

Pressure (PSI)	Period (μsec)	Frequency KHz	Δ P	Δ F	$\frac{\Delta F}{\Delta P}$
2.5	30.511697	32774.31603			
5.0	30.512184	32773.79292	2.5	.52311	.2092
7.5	30.512962	32772.95728	2.5	.83564	.3342
10	30.513841	32772.01320	2.5	.94408	.3776
12	30.514342	32771.47513	2.0	.53807	.2690
14	30.515131	32770.62779	2.0	.84734	.4236
16	30.515670	32770.04895	2.0	.57884	.2894
18	30.516672	32768.97297	2.0	1.07548	.5377
20	30.517467	32768.11932	2.0	.85365	.4268
22	30.518248	32767.28074	2.0	.83858	.4192
24	30.518981	32766.49374	2.0	.78700	.3935
26	30.519783	32765.63270	2.0	.86104	.4305
28	30.520500	32764.86296	2.0	.76974	.3848
30	30.521549	32763.73686	2.0	1.12610	.5630
32	30.522334	32762.89421	2.0	.84265	.4213
34	30.523220	32761.94320	2.0	.95101	.4755
36	30.524163	32760.93107	2.0	1.01213	.5060
38	30.525114	32759.91041	2.0	1.02066	.5103
40	30.526459	32758.46701	2.0	1.44340	.7217
42	30.526992	32757.89505	2.0	.57196	.2859
44	30.527970	32756.84561	2.0	1.04944	.5247
46	30.528805	32755.94967	2.0	.89594	.4479
48	30.532014	32755.16026	2.0	.78941	.3947

Bibliography

1. Brown, S.L. and S. Harris, Rev. Sci. Inst., vol 2, pg 180-183, Mar. 1931.
2. Roberts, E.A. and P. Goldsmith: Electrical Engineering, Vol 70, pg 776-780, Sept. 1951
3. Pacey, D.J., Vacuum, vol. 15, No 3, pg 95-98, Mar. 1965.

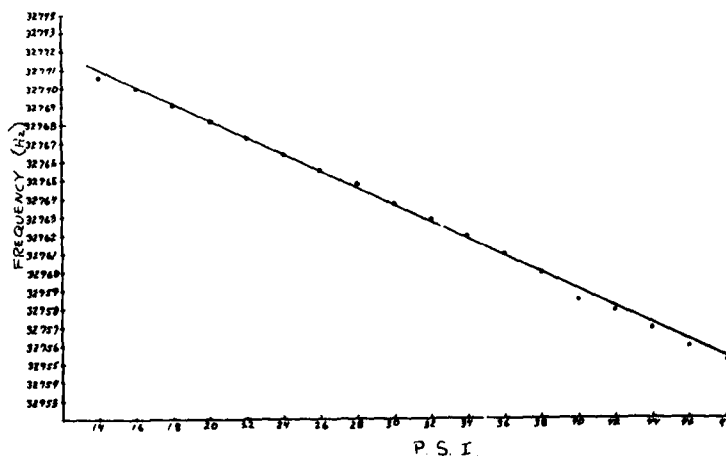


FIGURE A

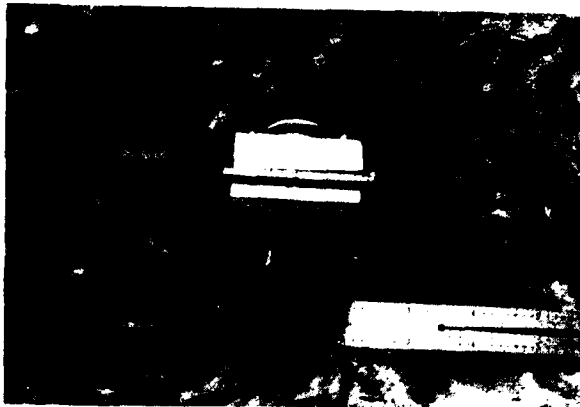


FIGURE B

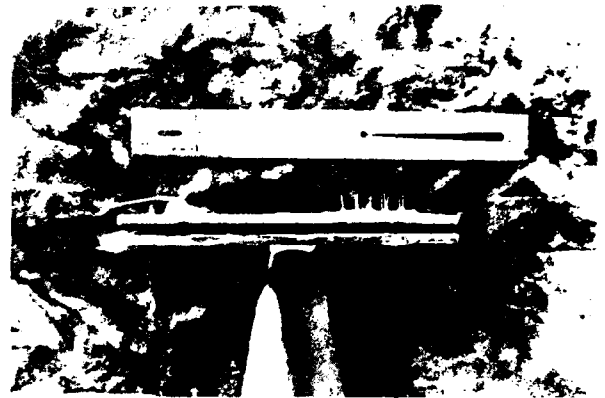


FIGURE C

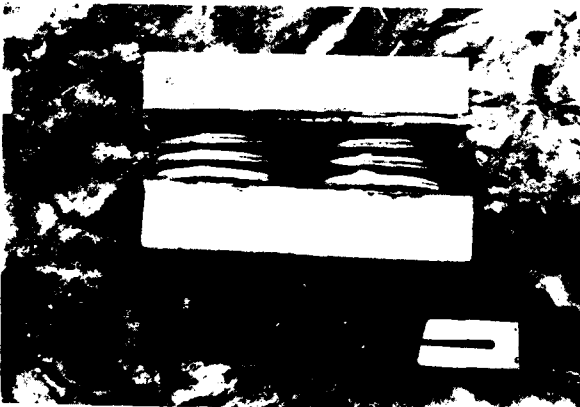


FIGURE D

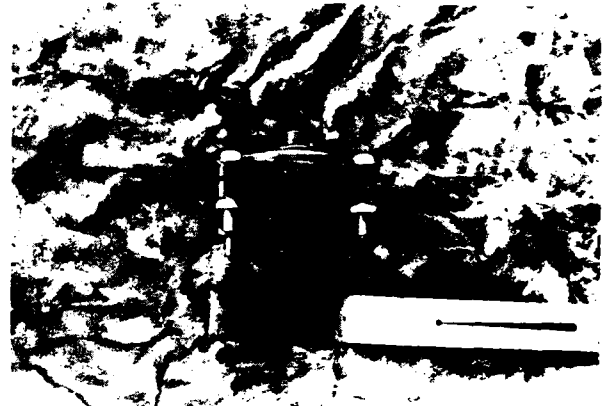


FIGURE E

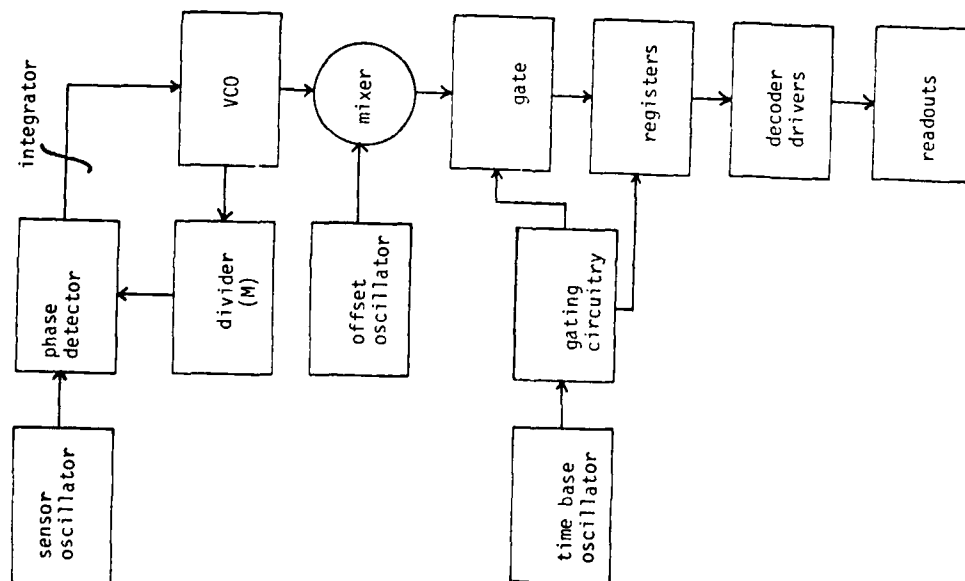


FIGURE F

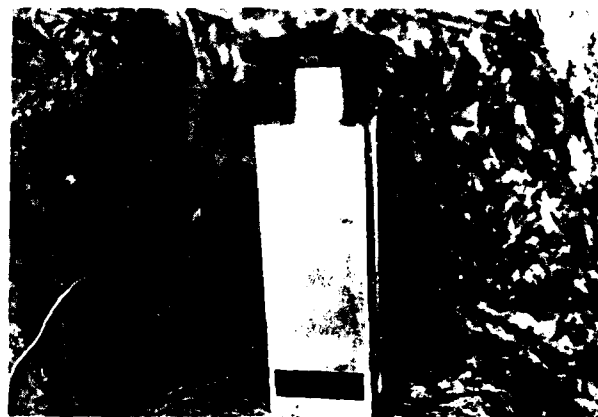


FIGURE G

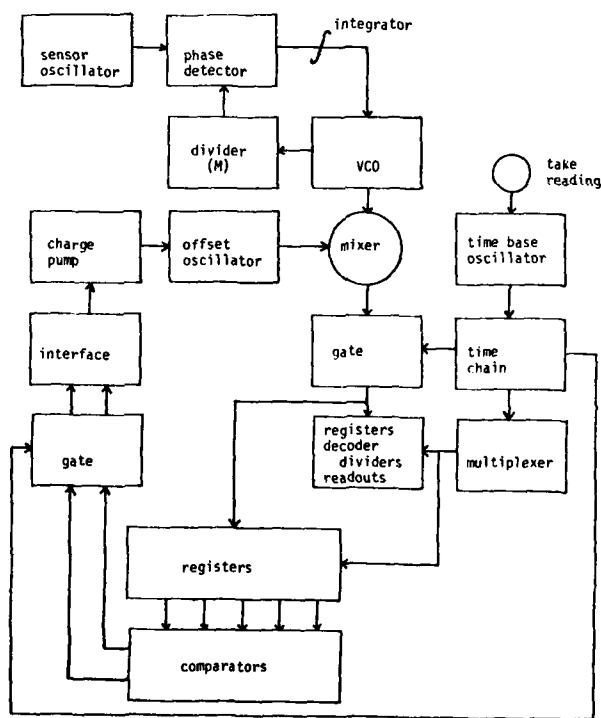


FIGURE H

TABLE 2

FORCE (grams)	PERIOD (μ sec)	ΔP (μ sec)
0	30.50817	
50	30.50819	
100	30.50822	.00002
150	30.50824	.00003
200	30.50828	.00002
250	30.50831	.00004
300	30.50834	.00003
350	30.50838	.00003
400	30.50841	.00004
450	30.50845	.00003
500	30.50849	.00004
550	30.50853	.00004
600	30.50856	.00003
650	30.50859	.00003
700	30.50863	.00004
750	30.50867	.00004
800	30.50871	.00004
850	30.50875	.00004
900	30.50878	.00003
950	30.50882	.00004
1000	30.50884	.00004
1050	30.50887	.00002
1100	30.50891	.00003
1150	30.50894	.00004
1200	30.50899	.00003
1250	30.50903	.00005
1300	30.50907	.00004
1350	30.50910	.00004
1400	30.50914	.00003
1450	30.50917	.00004
1500	30.50921	.00003
1550	30.50924	.00004
1600	30.50928	.00003
1650	30.50932	.00004
1700	30.50936	.00004
1750	30.50939	.00003
1800	30.50943	.00004
1850	30.50947	.00004
1900	30.50951	.00004
1950	30.50954	.00003
2000	30.50958	.00004

TABLE 2 (cont)

FORCE (grams)	PERIOD (μ sec)	ΔP (μ sec)
2050	30.50962	.00004
2100	30.50966	.00004
2150	30.50970	.00004
2200	30.50974	.00004
2250	30.50978	.00004
2300	30.50982	.00004
2350	30.50986	.00004
2400	30.50990	.00004
2450	30.50993	.00003
2500	30.50996	.00003
2550	30.51001	.00005
2600	30.51005	.00004
2650	30.51009	.00004
2700	30.51014	.00005
2750	30.51018	.00004
2800	30.51019	.00001
2850	30.51023	.00004
2900	30.51026	.00003
2950	30.51030	.00004
3000	30.51034	.00004
3050	30.51037	.00003
3100	30.51040	.00003
3150	30.51045	.00005
3200	30.51049	.00004
3250	30.51053	.00004
3300	30.51057	.00004
3350	30.51060	.00003
3400	30.51063	.00003
3450	30.51067	.00004
3500	30.51070	.00003
3550	30.51074	.00004
3600	30.51077	.00003
3650	30.51082	.00005
3700	30.51085	.00003
3750	30.51089	.00004
3800	30.51092	.00003
3850	30.51096	.00004
3900	30.51099	.00003
3950	30.51103	.00004
4000	30.51108	.00005
4050	30.51112	.00004
4100	30.51117	.00005
4150	30.51120	.00003
4200	30.51124	.00004

TABLE 2 (cont)

FORCE (grams)	PERIOD (μ sec)	ΔP (μ sec)
4250	30.51128	.00004
4300	30.51131	.00003
4350	30.51134	.00003
4400	30.51137	.00004
4450	30.51142	.00005
4500	30.51147	.00005
4550	30.51150	.00003
4600	30.51154	.00004
4650	30.51157	.00003
4700	30.51161	.00004
4750	30.51165	.00004
4800	30.51170	.00005
4850	30.51173	.00003
4900	30.51176	.00003
4950	30.51180	.00004
5000	30.51184	.00004
5050	30.51189	.00005
5100	30.51193	.00004
5150	30.51195	.00002
5200	30.51199	.00004
5250	30.51203	.00004
5300	30.51207	.00004
5350	30.51210	.00003
5400	30.51214	.00004
5450	30.51218	.00004
5500	30.51222	.00004
5550	30.51226	.00004
5600	30.51229	.00003
5650	30.51233	.00004
5700	30.51237	.00004
5750	30.51240	.00003
5800	30.51245	.00005
5850	30.51249	.00004
5900	30.51252	.00003
5950	30.51256	.00004
6000	30.51260	.00004

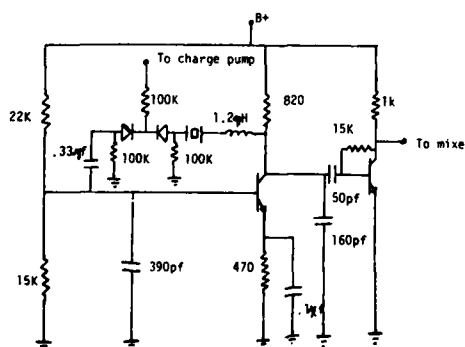
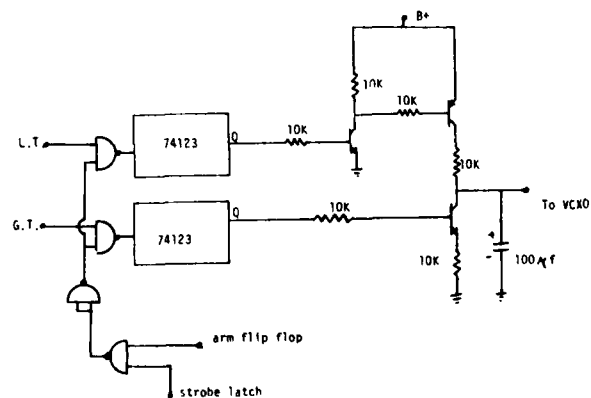
OFFSET VCXO
FIGURE KLOGIC INTERFACE AND CHARGE PUMP
FIGURE J



Figure L1



FIGURE L3



FIGURE L2

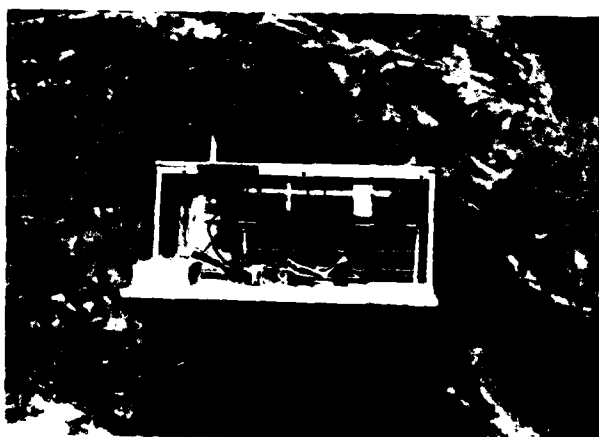


FIGURE L4

METHODS FOR PRODUCTION SCREENING FOR ANOMALOUS RESPONSES IN QUARTZ CRYSTALS INTENDED FOR HIGH RELIABILITY APPLICATIONS

Patrick F. Godwin, Jr. and George L. Snider,
TRW Defense and Space Systems Group
Redondo Beach, California

Introduction

Many prime contractors and subcontractors have become aware of anomalous performance in crystal oscillators and crystal filters in the past three years. Frequently these anomalies have gone undetected in flight hardware in tests up to system integration and launch vehicle/payload levels of testing. The most frequent symptoms encountered in crystal filters are phase and amplitude discontinuities (e.g., signal drop-out) over a limited temperature range. In crystal oscillators, the most pronounced effect is a frequency jump which can vary from as small as one part per 100 million (1E8) to 20 parts per million (PPM) occurring over a limited temperature range.

Anomalies appear to be a random occurrence in crystal production seemingly independent of supplier and may occur in 10 to 30 percent of production crystals. Anomalies are observed in temperature ranges as narrow as 0.1°C and as wide as 20°C (see Figures 1, 2, and 3 for examples from three crystal suppliers). The point in temperature where anomalies are encountered tends to be random within any given crystal design (although systematic anomalies have been observed in some instances). Anomalies can be induced by poor design of oscillator or filter circuits or by poor definition of crystal requirements (e.g., too high a crystal drive level or operation at a drive level not specified to the supplier). Pseudo anomalies can also be introduced by thermal transients which result from conditions that are unreal for the proposed crystal hardware application. Pseudo anomalies result from test fixture temperature slew rates which are far in excess of temperature slew rates expected in an operational environment. Current supplier test methods permit a significant number of crystals with anomalies to escape detection.

The objectives of this paper are to make the general high reliability part user aware of the problem and to define test requirements and test methods which can be used to provide an effective screening of production crystals to preclude the acceptance of units subject to discontinuous responses into hardware.

Test Requirements

The requirements of a crystal test method adequate to assure detection of anomalies include provision for:

1. Operation of the crystal at the energy level (drive level) required by circuit application.
2. Identification and separation of resonant frequency shifts due to apparent crystal mass changes as opposed to frequency changes in test circuits induced by activity changes.
3. Identification of resistance (activity) changes. (The effect of resistance change is a function of parameter sensitivity of application circuit.)
4. Identification of temperature dependence with a resolution of 0.05°C over any required test temperature range.

5. Identification of temperature slew rate effects to isolate steady anomalies from pseudo (transient) induced effects.

Test Methods

The test requirements stated above can be met by treating the crystal under test as a reference in a phase lock loop. Phase lock loop parameters can be chosen such that crystal operation may be maintained at resonance relatively independent of crystal resistance and operating drive level. This permits isolation of the effects of each parameter in diagnostic activity and production screening under conditions compatible with the intended application.

Three measurement implementations are described in sufficient detail to permit each implementation to be emulated according to the needs of potential users. The three basic implementations are:

1. Combination of standard laboratory RF test equipment including synthesizers, vector voltmeters, double balanced mixers, amplifiers, counters, etc.
2. Customized test system using off-the-shelf ECL, TTL, and linear integrated circuits with optional analog X/Y plotter readout or microprocessor-compatible 8-bit latched digital outputs. This approach can be economically realized as a multichannel system.
3. Computer controlled multichannel system utilizing standard bus programmable RF test instrumentation.

The theory of the proposed test method will be explained through a detailed presentation of the customized test system implementation.

Figure 4 is a general block diagram of a phase lock loop which is based on the use of a crystal as a "zero" phase reference at the desired operating frequency. The basic loop consists of a voltage-controlled oscillator, phase detector, and feedback amplifier. A double balanced mixer is used in conjunction with a 90 degree "lossy" delay line to act as a linear phase detector. The crystal under test acts like a series bandpass filter in a test circuit configuration generally referred to as a π circuit configuration. The dynamics of the control loop formed results in frequency lockup at the zero phase frequency of the crystal.

Phase change of the lossy delay can be on the order of 1 degree per megacycle at 85 megacycles. Phase change of the crystal will be on the order of 50 degrees for a frequency change of 10 parts per million at 85 megacycles. On a relative scale the lossy delay line behaves as a constant 90 degree reference. These relationships are demonstrated for typical values in Figures 5a, 5b, and 6.

From Figure 6 the transfer characteristic of the " π " (zero phase reference of Figure 7) can be approximated from the computed phase versus frequency response shown.

$$\frac{\Delta \phi}{\Delta F} = \frac{50^\circ (\pi/180^\circ) \text{ radians}}{(10 \times 10^{-6})(85 \times 10^6)(2\pi) \text{ radians/second}}$$

$$\frac{\Delta \phi}{\Delta F} = 1.63 \times 10^{-4} \frac{\text{radians}}{\text{radians/second}} \quad (1)$$

A rough prediction of phase detector sensitivity may be made by multiplying L and R inputs to the balanced mixer (see Figure 7) (assuming inputs in phase) and dividing by 90 degrees ($\pi/2$ radians). This ignores mixer losses and assumes square wave inputs. Emitter-coupled logic swings based on the loads shown in the circuit configuration of Figure 7 vary from 0.4 volts peak-to-peak to 0.7 volts peak-to-peak as a function of test frequency. These logic swings follow the zero crossings of the π output signal and phase reference input signal. The phase detector sensitivity becomes:

$$\frac{0.7 \text{ v} \times 0.7 \text{ v}}{90^\circ} = \frac{544 \text{ mv}}{1^\circ \text{ Phase}}$$

$$= \frac{0.7(0.7)}{\pi/2} = 0.3119 \frac{\text{volts}}{\text{radian}} = K'd \quad (2)$$

The amplifier shown in Figure 7 as a level translator and integrator has a transfer function response of $100/[1 + j\omega(C_1)]$. Note that this stage provides a dc gain of 100 and a single pole filter (integration) response in the feedback loop. If the amplifier response is lumped with the detector response, the overall detector response becomes:

$$K'd = 100 \frac{\text{volts}}{\text{volt}} \left(0.3119 \frac{\text{volts}}{\text{radian}} \right) = 31.19 \frac{\text{volts}}{\text{radian}} \quad (3)$$

at loop phase error frequencies near dc.

The range of the voltage-controlled oscillator requires restriction by broad range tuning and bias adjustment to preclude lockup during test at undesired crystal resonance frequencies. Crystals can be induced to oscillate at any of their resonant modes in a predictable manner where this is deemed desirable for investigative purposes. These conditions are predictable and are under simple manual or automatic control by the operator. For the normal case (i.e., test of crystal at its desired operating mode), the upper frequency limit of the voltage controlled oscillator is simply controlled to a value of frequency below that of the first spur (undesired response) above the main mode. Note that oscillator output amplitude is unaffected by frequency change induced by voltage input to the tuning section of the VCO (over a range of interest in the test of any given crystal which is typically much less than 100 parts per million).

The transfer function characteristic of the voltage-controlled oscillator (VCO) is simply estimated as follows from the tank circuit parameters illustrated in Figure 7 and the typical varactor diode pair capacitance versus voltage characteristic shown in Figure 8.

$$F_{\text{Test}} \approx 85 \text{ megahertz} = \frac{1}{2\pi\sqrt{L_1 C_1}} \quad (4)$$

Let

$$L_1 = 0.5 \mu\text{H}$$

and

$$C_1 = 7.01 \text{ pF}$$

Assume

$$C_{11} = C_{\text{Fixed}} + C_{\text{Diode}}$$

Referring to Figure 8, at 15 volts bias, $C_D \approx 5.15 \text{ pF}$

$$C_{\text{Fixed}} = C_1 - C_D = 7.01 - 5.15 = 1.86 \text{ pF}$$

If C_D is biased at 10 volts, $C_D \approx 6.0 \text{ pF}$ from Figure 8. With the feedback loop open (i.e., Switch #1 in Figure 7 shorted) and the varactor biased at 10 volts, the VCO output frequency is approximately

$$F = \frac{1}{2\pi\sqrt{L_1 C_1}} = \frac{1}{2\pi\sqrt{(0.5 \times 10^{-6})(1.86 \times 10^{-12})}} \quad (5)$$

$$F \approx 79.57 \text{ MHz at 10 volts bias}$$

and

$$C_{\text{Fixed}} = 1.86 \text{ pF}$$

Note that a 1 pF change results in a delta frequency of $85 \times 10^6 - 79.57 \times 10^6 = 6.03 \text{ mhz}$ or an average scale factor of 6.03 megahertz per pf. From figure 8 the capacitance change from a bias voltage of 10 volts to 15 volts (saturation of the operational amplifier) results in a capacitance change of 1 pf or a scale factor of 1 pf/5 volts $\approx 0.2 \text{ pf/volt}$. The total transfer function " K_O " of the VCO for the circuit values chosen becomes:

$$K_O = \frac{1 \text{ pF}}{5 \text{ v}} \times \frac{6.03 \text{ MHz}}{1 \text{ pF}} = \frac{1.206 \text{ MHz}}{\text{volt}}$$

$$= \frac{7.577 \times 10^6 \text{ radians/sec}}{1 \text{ volt}} \quad (6)$$

The control loop dynamics for the phase lock loop shown in Figures 4 and 7 may be summarized in the open loop incremental model shown in Figure 9. From Figure 9 the total open loop gain at error frequencies near dc is

$$20 \log_{10} \left[K'd \times \frac{100}{1 - j\omega(1)} \times K_O \times \frac{\Delta \phi}{\Delta F} \right] \approx 91 \text{ dB} \quad (7)$$

The loop has a simple 1 pole response with a 3-dB frequency of 0.159 hertz.

It is useful at this point to summarize the conditions for operation of the crystal under test through the example of 85 megahertz crystal parameters illustrated. A typical test procedure would involve selection of L_1 , C_1 and tuning of the VCO (with the feedback loop open, i.e., switch 1 shorted to ground) with the output bias of the amplifier (input to varactor) set to +10 volts dc. The VCO output frequency under these conditions would be set to about

79 megahertz. An "a priori" test (swept response of the crystal) would indicate crystal spurs (inharmonic responses) in the neighborhood of less than 0.5 percent above the main mode (e.g., 100 kHz to 300 kHz above the main mode). The fixed bias can be driven to +15 volts from this tuning point to verify that the highest frequency of the VCO will be less than the frequency corresponding to the first spur. With the varactor bias set to +10 volts, open the shorting switch and the loop will lock up on the main mode response of the crystal. The loop is unconditionally stable over a wide range of operating conditions. Crystal drive level may be simply computed from the input voltage to the π recognizing that the crystal is held at series resonance. The writers chose to use a separate hybrid wide band amplifier and power splitter to allow for a wide range of adjustment of crystal drive levels. Considerable latitude in location of the π (crystal under test) is allowed with this additional buffering. In addition, the values of circuit π resistors may be varied to set desired drive level. Again, the reader is reminded that the objective is to test for discontinuity as opposed to absolute value measurement and concern for absolute value correlation of data. The VCO shown has a built-in amplifier option which may be useful to the user with a limited screening requirement (this is a tuned amplifier stage).

What does the reader do if other resistor values are chosen for the π or if some idea of correlation in measurement is desired? The writers in this paper have included an SR-52 program which accepts any combination of π component values and predicts gain response, phase response, and crystal drive level in Table I. The π resistor network was scaled to provide a minimum input signal of 5 mV peak (3.5 mV rms) to the input of the squaring amplifier/comparator which forms an integral part of the phase detector. For a 200 ohm crystal and the π values shown, a minimum input to the should be 163 mV rms.

Initial alignment is simply accomplished upon completion of circuit fabrication by removing the capacitors in the lossy delay line and adjusting coax line lengths to peak output of the detector. This alignment should hold for any test frequency. The lossy delay line is simply calibrated by adjusting the capacitors to yield minimum detector output with the feedback loop open. An SR-52 program is included in Table II which can be used to compute lossy delay unit select-in-test capacitor values which will yield 90° phase shift at the test frequency. (For equal value select-in-test capacitor values and resistors as shown this becomes a one line computation as shown in Figure 5a.)

The preceding presentation defines a circuit which establishes the crystal under test at a desired mode of operation (main mode or spur) in an unconditionally stable phase lock loop. The frequency output of the voltage controlled oscillator is dominated (or controlled) by the series resonance frequency property of the crystal under test. Crystal frequency can be simply monitored by use of a linear counter type discriminator illustrated in Figures 4 and 10. Crystal resistance can be monitored by detecting output voltage from the π circuit shown in Figure 7 through a coherent detector described in Figure 11. Note that with the loop in phase lock the steady state RF voltage output of the π is determined by the voltage divider formed by the fixed resistors shown and the crystal replaced by its equivalent series resistance. This π output is coherently detected by the circuit shown in Figure 11 in the following manner: (1) The zero crossings of the π output are detected and amplified by one-half of the dual 687 ECL comparator and one-fourth of the 1692 ECL line receiver and fed as a constant amplitude input to

the "L" input of the double balanced mixer. The propagation delay specified for the 687 and 1692 is on the order of (6 to 8) nanoseconds. The propagation delay from 687/MC1692 resulting in a phase lag appears as a phase lead at Pin #7 (steady state due to inverting in the amplifier). (2) The output of the π is coupled through a 50 ohm cable of sufficient length to match this delay (approximately 10 inches of RG 316). The cable is simply tuned to match the ECL part delays by peaking the dc output of the double balanced mixer at the "I" output. This adjustment, when completed in initial fixture alignment, remains frequency independent. The amplifier shown in Figure 11 provides for scale factor and voltage offset adjustment. Voltage output polarity may be reversed by interchanging pin 2 and pin 3 of the OP-07 amplifier in Figure 11. The voltage offset adjustment may be used to null the detector output at some minimal value of crystal resistance in order to improve detector sensitivity to crystal resistance changes (allows display by user with higher detector gain without "off" scale performance of recorder).

Figures 12 and 13 illustrate a temperature monitor circuit which has been found to track within a 1 degree scale factor error at -55°C and +100°C temperature extremes. A slightly more simple circuit using a National Semiconductor LX5700 integrated circuit temperature sensor in a TO-46 package exhibited less than a 2 degree centigrade error under similar test conditions.

Note from Figure 10 that delta frequency information is directly available in analog or digital format. The R_x plug-in resistor sets the counting interval (T1) for the difference frequency counter. The counter reset pulse (T2) is available to generate an interrupt request to a microprocessor, computer/calculator, or minicomputer to indicate that the latched dc outputs of the 93L08 latch are indicating an 8 bit parallel digital formatted valid value of delta frequency. An 8 bit parallel-to-serial shift register would be required to acquire the test data in a serial digital format.

Figure 14 describes a simple TTL circuit which can be used to translate temperature and crystal resistance data from analog to digital format. The digital data can be supplied in 8 bit parallel or serial format from the circuit in Figure 14 (as shown) to be compatible with a microprocessor, computer/calculator (e.g., HP9825) or minicomputer interface. An 8 bit serial output is available at the serial output of the 2502 (25L02) successive approximation register (SAR) during the first 8 clock intervals following a data request (start) signal from a microprocessor. Data is valid in 8 bit parallel digital format in the clock interval following the end of conversion signal from the 2502 (25L02) SAR. CMOS integrated circuits are becoming available (e.g. AD7570) to supply the total analog-to-digital converter function (typically CMOS circuits are less tolerant of handling abuse in test circuit fabrication but are otherwise desirable).

If the reader were to compute output voltage of the crystal as a function of crystal resistance with π phase shift nulled to zero, it would be apparent that output voltage is not a linear function of crystal resistance. The input voltage to the test circuit is independent of crystal equivalent resistance (i.e., stays constant). Therefore, for a fixed amplitude input at crystal resonance a calibration curve can be generated for use in a direct analog display. A simple conversion table can be generated digitally from a calibration curve in a number of ways. One simple

implementation within a computer is to use a look-up table. An alternate scheme would use a PROM with the digital encoder output of the applied to the address lines of the PROM and the data in the address location of the PROM reflecting the true value of resistance derived from calibration.

Summing Up Custom System Test Method

Recurring tuning requirements include the LC tank for the ECL oscillator and a capacitor for the 'LOSSY' delay line. Each selection can be based on a 1-line computation for any given crystal type. The parts required are generally available to anyone interested in their procurement. For users with a single crystal type to be screened, certain economies are obvious. For example, the tuned amplifier stage in the ECL oscillator can be used to increase signal input power in lieu of use of the thin film wide band hybrid amplifier (trading-off one more tuning step for an additional component). A simple method for making the crystal lock up in an unconditionally stable loop is demonstrated. Control loop frequencies which introduce additional phase shifts are avoided (e.g., frequency range where compliance of crystal is frequency dependent (generally over 20 kilohertz) as evidenced by many voltage controlled crystal oscillator input frequency constraints). Settling time (for the case where a crystal is maintained in a closed loop as shown in Figure 7) is easily verified empirically for any given crystal. Settling time can be a significant factor involved in determining the length of time required to perform a test while maintaining temperature continuous in its variation through the test (e.g., temperature steps less than 0.1 degrees centigrade). By simply removing the shunt across the 20 pF capacitor in the π (see Figure 7) loop response time can be visually monitored on a recorder. The loop time constant set by the integrator is 1 second in the example shown. This translates to setting an acceptance limit to 70 percent of the specification value if it is desired to sweep temperature at a rate of 0.05 degrees per second and still screen anomalous responses within a specified acceptance limit. The L/R time constant for the sample 85 MHz crystal shown is on the order of 6.5×10^{-3} henries/5 ohms \approx 1.3 milliseconds. The L/R time constant for a 5 MHz frequency standard crystal could well be on the order of 25 henries/5 ohms \approx 5 seconds. In one instance the time constant of the integrator dominates settling time response. In the other case it does not. In either case, the empirical test described is an easy way to bound measurement response time requirements for a given class of crystals.

Generalizations Which Can Be Applied To Computer Test Method

Each of the basic elements in the custom system described can be replaced by certified, commercially available, computer bus oriented units of test equipment. Many frequency synthesizers can be operated in a voltage controlled oscillator (VCO) mode. A tracking receiver or vector voltmeter can be utilized as a phase detector and RF voltmeter and finally, period type counters such as the HP5345 allow for accurate direct frequency readout.

The same loop stability constraints apply when measuring crystal parameters under computer control as apply for the custom system. Minimum measurement settling time will prevail when the crystal under test is maintained at resonance in a closed loop circuit such as that shown in Figure 7. If a computer is used to

test a number of crystals which are concurrently being exposed to temperature change, each crystal under test can be supported in a separate closed loop circuit such as that shown in Figure 7 to minimize test time as determined by circuit settling time response.

Some suppliers have found it desirable to check a number of crystals concurrently by use of a turret type mechanical switch handling a load of crystals or by use of an electronic RF switching array. The principal impediment to continuous temperature response type testing with these implementations is the response time associated with measuring each crystal response from no excitation to switch closure (selection of crystal under test). The use of predictive techniques to enhance measurement response has now ~~yet~~ been implemented in computer software by at least one supplier.

Each measurement can start with a crystal input frequency which is relatively wide of the mark for crystal resonance at the test temperature. Over the relatively wide span in temperature between turning points crystal frequency change is quite predictable for a given lot and specified turning point tolerance. Except when an anomaly is impacting response of the crystal, successive measurements at contiguous temperatures could begin with test frequency programmed virtually "on the mark" with respect to frequency of "0" phase response. This approach to minimizing measurement response time requires a frequency source which can be modulated (switched) at a rate which is fast relative to crystal response times. This approach also requires more of the test computer in memory with faster access times and computational capacity. It also requires more sophistication in making the measurement time "adaptive" in proportion to the null error sensed and rate of closure to a null with each measurement.

Some current computer controlled test systems tend to avoid "continuous" measurements of frequency versus temperature and resistance versus temperature. Temperature is allowed to stabilize on the order of 15 minutes prior to each set of measurements. Emphasis is placed on absolute value data and no special effort is evident in reducing measurement time to improve "throughput" with respect to screening production quantities of parts. Continuous temperature testing by high reliability crystal suppliers tends to be defined as measurement at 3°F intervals for the "turret type" computer controlled test system. In the writer's opinion this is a grudging recognition that this level screen would have detected crystals with anomalies that caused holes in the response of crystal filters intended for high reliability UHF communications receivers in satellite communication's systems.

Crystal product suppliers for outstanding procurements of parts for the DSCSII Satellite Communication System have accepted a requirement to screen crystals over temperature with test points at 0.05°C intervals being accepted as a continuous measurement. Figure 15 illustrates the effect of crystals with very narrow temperature anomalies which were screened out by one oscillator supplier by tests conducted at an oscillator level of assembly. Note that two out of three crystal responses shown in Figure 15 exhibited anomalies on the order of 1°C. The oscillator responses shown were subsequently verified to be a result of a narrow band of temperature anomalies in the crystals used in the oscillators. Test times for the examples shown in Figure 15 were 19 hours. Minimization of test time was not a significant factor considered in establishment of this particular test.

Generalizations Applicable To Test Method Using
Current Standard RF Test Equipment

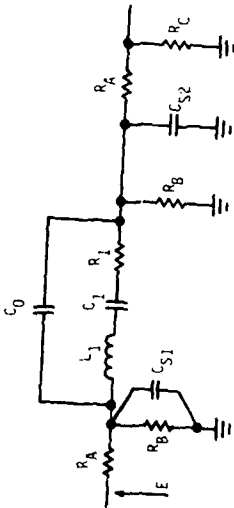
The same control loop dynamics apply as for the custom circuit approach described. It has been empirically discovered that certain frequency synthesizer and vector voltmeter combinations are dc compatible to the extent that they can function as a voltage controlled oscillator and phase detector with a simple capacitor used to form a passive integrator (developing a loop with a single pole response) (e.g., Figures 16, 17, with 18 and 19.) The synthesizer must have a low FM present in its output to prevent "washing out" narrow frequency band effects. The vector voltmeter should be used with a 90 degree offset to

function in the manner described for the customized circuit (i.e., present 0 dc output for 0° phase shift across the crystal in the π). Control loop dynamics can be verified in the same manner empirically as described for the customized test circuit method (e.g., removing the shunt across the 20 picofarads shown in the π and observing settling time response on an analog recorder). Unless the frequency synthesizer sweep range in a V_{CO} mode is limited in the same manner as described for the customized circuit, the test technician may require several trials to cause loop closure following the "main" (desired) mode response as opposed to loop lock on a spur (inharmonic) response. Lack of limiting V_{CO} range can result in the loop "jumping" modes in response to a transient disturbance (e.g., power supply interruption or noise).

Table I. Transfer Functions for Crystal "π" Test Circuit as a Function of Circuit Parameters

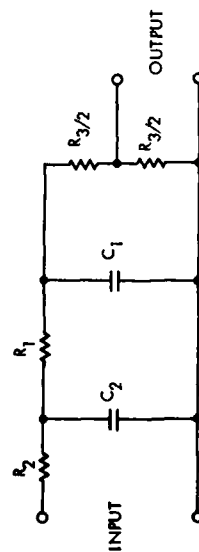
[illegible]

SR52 USER PROCEDURE:



1. COMPUTER $R_{AB} = R_A \cdot R_B$ AND $P_{ABC} = \frac{(P_A + P_C) \cdot R_B}{R_A + R_B + R_C}$
2. LOAD PROGRAM CARD SIZE A AND SIDE B
3. LOAD CIRCUIT PARAMETERS IN DATA REGISTERS AS FOLLOWS: R_{ABC} STO 19, P_{ABC} STO 19, R_C STO 17, L STO 15, C STO 14, C_2 STO 13, C_1 STO 12, $\omega = 2\pi F$ STO 11, C_0 STO 10
4. PRESS E AND READ CIRCUIT PHASE SHIFT
5. PRESS RUN AND READ THEVENIN CIRCUIT GAIN. FOR TOTAL GAIN MULTIPLY BY $\frac{R_B}{R_A + R_B} \cdot \frac{R_C}{R_A + R_C}$
6. PRESS RCL, 0, 0 AND TOTAL THEVENIN CIRCUIT IMPEDANCE WILL BE DISPLAYED
7. PRESS SECOND, X AND TOTAL THEVENIN CIRCUIT ADMITTANCE DISPLAYED = Y TOTAL
8. COMPUTE I TOTAL = E TIMES $\frac{R_B \cdot R_C}{R_A + R_B}$ TIMES Y TOTAL, COMPUTER DRIVE DRIVE LEVEL = (I TOTAL $2R_1$)

[illegible]

[illegible]

SR52 USER PROCEDURE;

1. ENTER PROGRAM SIDES A AND B
2. ENTER DATA FOR CIRCUIT MODEL IN DATA REGISTERS
 $F \rightarrow STO 14$
 $R_3 \rightarrow STO 15$
 $R_2 \rightarrow STO 16$
 $R_1 \rightarrow STO 17$
 $2C_2 \rightarrow STO 18$
 $2C_1 \rightarrow STO 19$
3. PRESS A TO EXECUTE PROGRAM. DISPLAY HALTS AT GAIN IN dB
4. PRESS RUN, DISPLAY HALTS AT TOTAL PHASE SHIFT
5. PRESS RCL 11
6. CHOOSE A NEW VALUE (S) FOR VARIABLES e.g., F AND REPEAT STEPS 2-5.

112	43	RCL	123	43	RCL	134	95	=	145	00	0	156	00	0	167	00	0	178	54)	189	01	1	200	00	0	211	43	RCL		
113	01	1	124	00	0	135	42	STO	146	00	0	157	09	9	163	00	0	179	22	INV	190	95	=	201	05	5	212	00	0		
114	08	8	125	04	4	136	00	0	147	20	*1/X	158	65	X	169	43	RCL	180	39	*P/R	191	43	RCL	202	55	+	213	08	8		
115	65	X	126	65	X	137	07	7	148	42	STO	159	53	(170	00	0	181	42	STO	192	00	0	203	02	2	214	75	-		
116	59	*π	127	53	(138	22	INV	149	00	0	160	43	RCL	171	09	9	182	01	1	193	09	9	204	95	-	215	43	RCL		
117	65	X	128	43	RCL	139	39	*P/R	150	09	9	161	00	0	172	65	X	183	00	0	194	55	+	205	107	10	216	01	1		
118	43	RCL	129	00	0	140	94	=	151	43	RCL	162	08	8	173	53	(184	43	RCL	195	43	RCL	206	65	X	217	00	0		
119	01	1	130	03	3	141	42	STO	152	01	1	163	33	COM	174	43	RCL	185	00	0	196	01	1	207	02	2	218	85	+		
120	04	4	131	32	SIN	142	00	0	153	06	6	164	54)	175	00	0	186	00	0	197	01	1	208	00	0	219	43	RCL		
121	85	+	132	54)	143	08	8	154	85	+	165	95	=	176	08	8	187	42	STO	198	65	X	209	95	-	220	00	0		
122	53	(133	54)	144	43	RCL	155	43	RCL	166	42	STO	177	32	SIN	188	01	1	199	43	RCL	210	81	HLT	221	06	6		
																										222	95	-			
																										223	81	HLT			

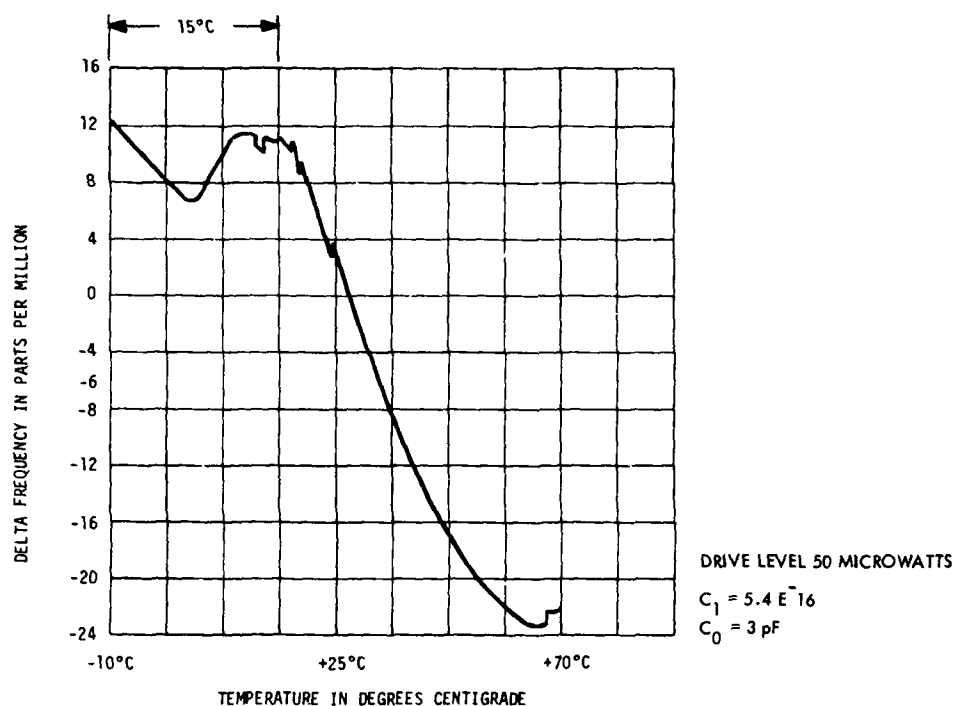
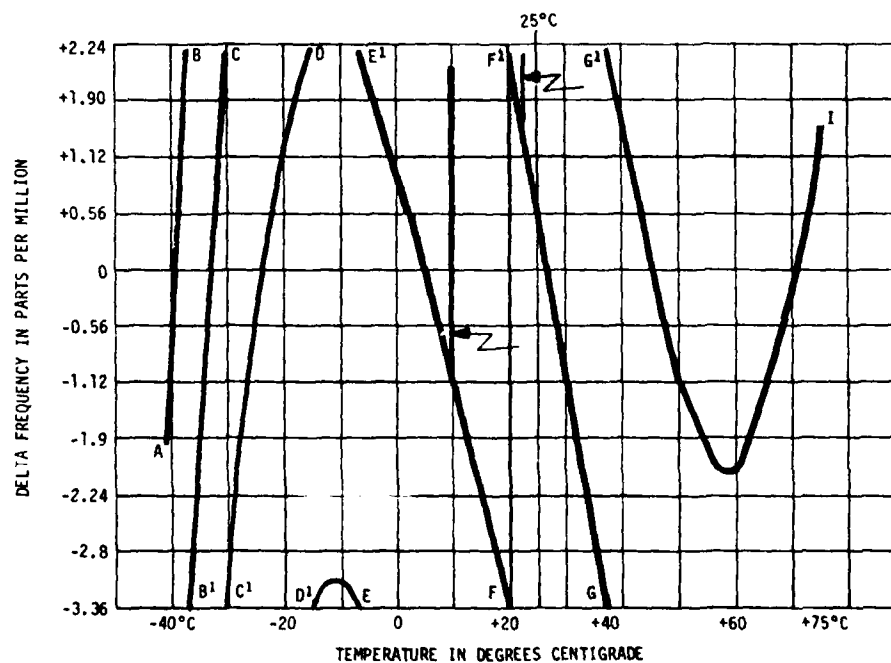


Figure 1. Frequency vs Temperature for a Crystal with a Wide Temperature Anomaly



NOTES:

1. 2 glitches (anomalies) at arrows as shown.
2. X/Y recorder zeroes as shown at B to B¹, C to C¹, D to D¹, E to E¹, F to F¹, G to G¹. The effect is to magnify Y scale by a factor of 6 over that required to fit the chart without a folding scale.

Figure 2. Frequency vs Temperature for a Third Overtone Crystal in a TO-5 Can With Two Narrow Temperature Range Frequency Anomalies

AD-A088 221

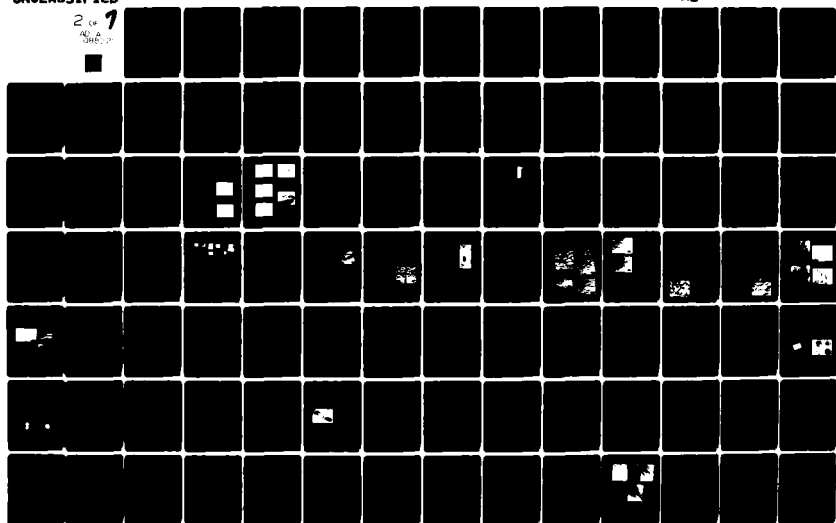
ARMY ELECTRONICS COMMAND FORT MONMOUTH NJ
PROCEEDINGS OF THE ANNUAL FREQUENCY CONTROL SYMPOSIUM (31ST) ME--ETC(U)
1977

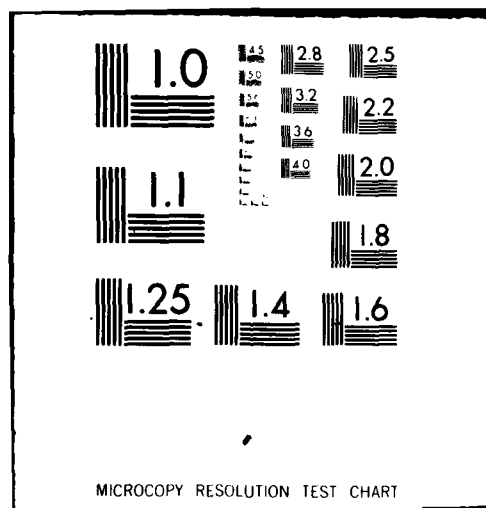
F/G 9/5

UNCLASSIFIED

NL

2 of 7





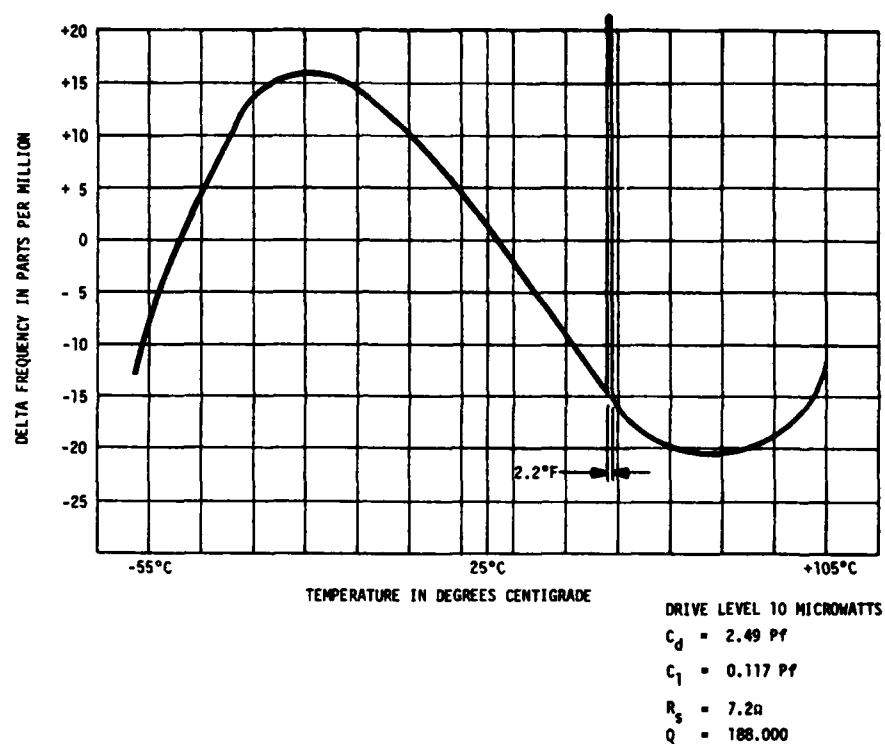


Figure 3. Frequency vs Temperature for a Crystal with a Narrow Frequency Anomaly

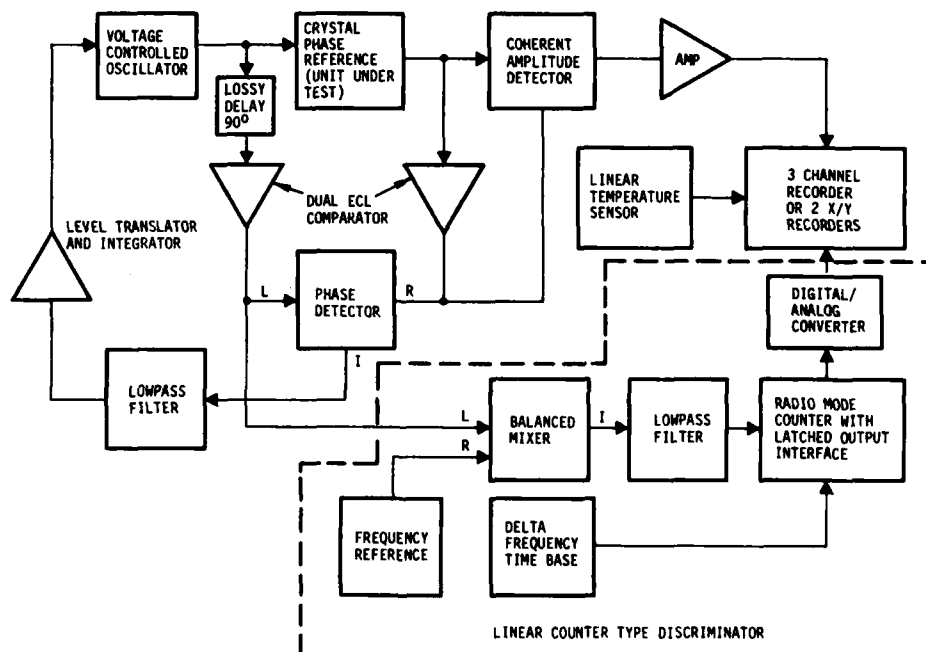
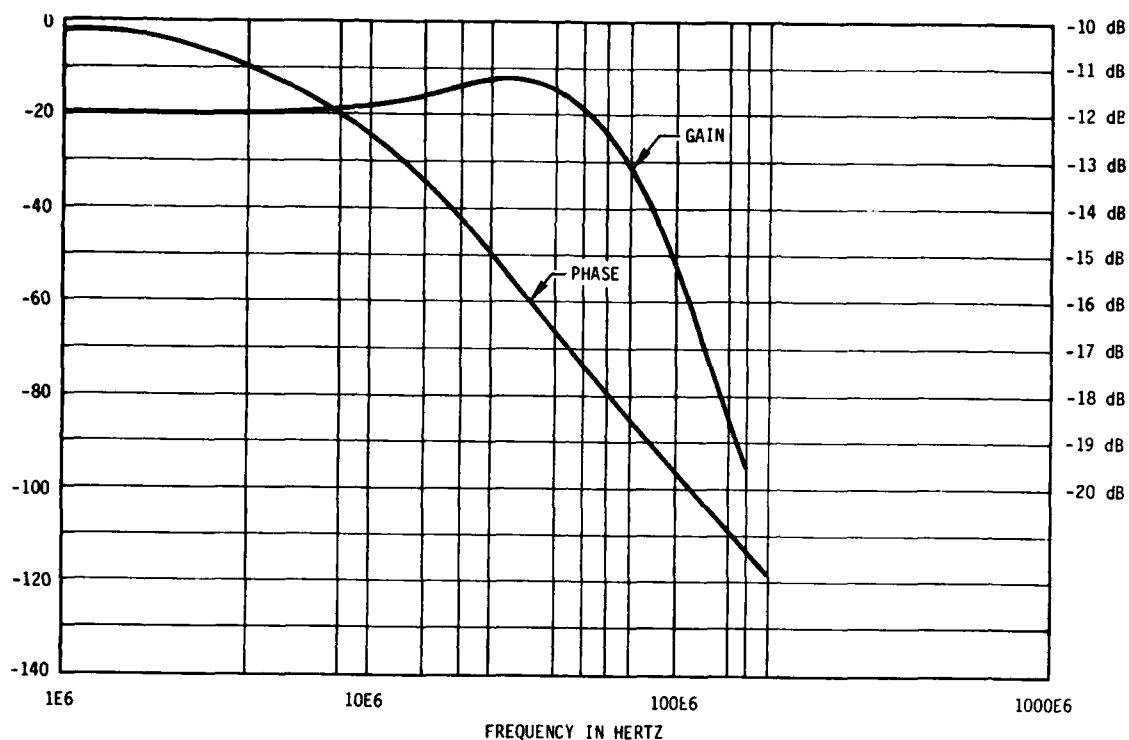


Figure 4. Phaselock Loop System for Detecting Anomalous Quartz Crystal Frequency and Activity Changes



NOTE

1. $C = 1/434.1F_0$
2. PHASE SHIFT = -90° AT F_0
3. INSERTION LOSS = 14 dB
4. VSWR = 1.4 at F_0

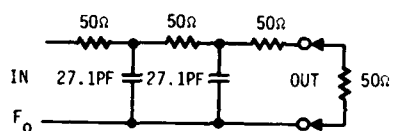


Figure 5a. Gain and Phase vs Frequency for Lossy Delay Circuit

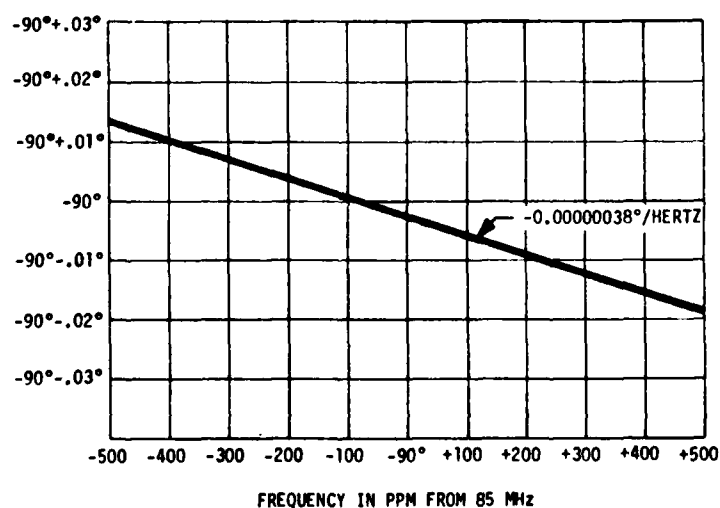
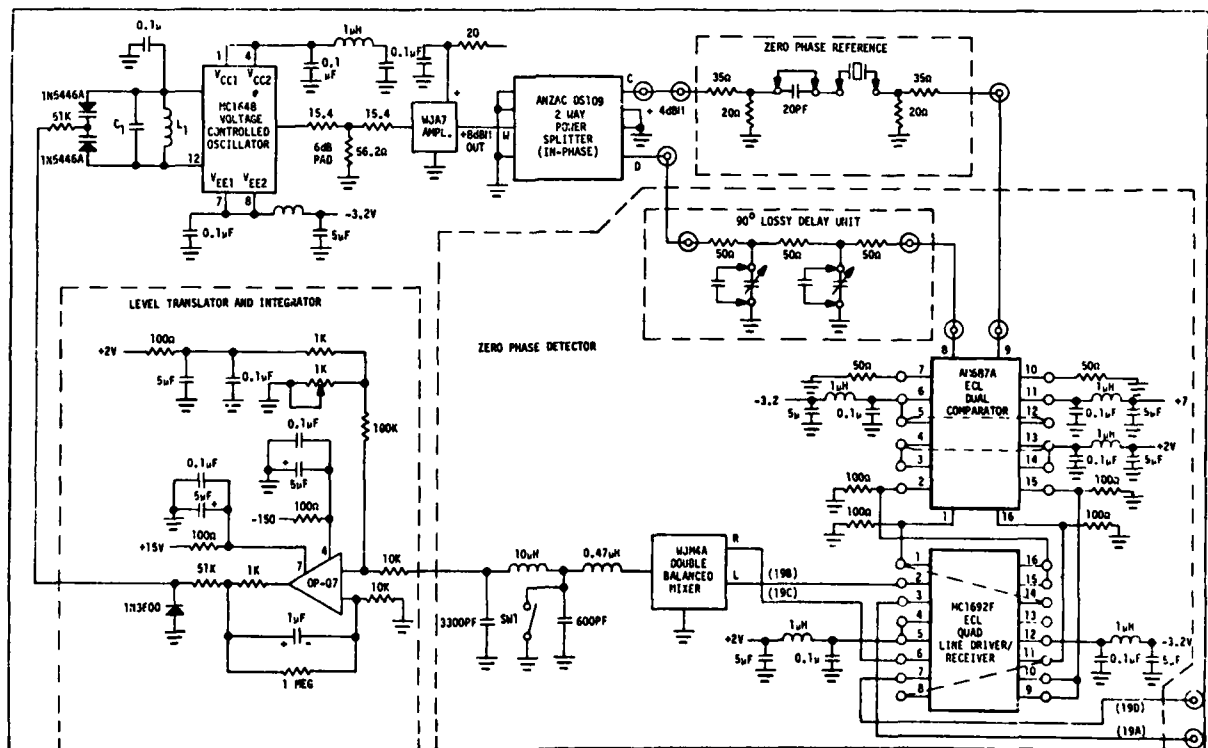
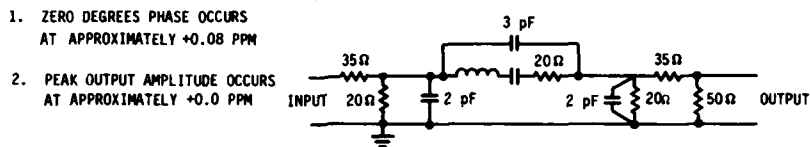


Figure 5b. Lossy Delay Phase vs Frequency



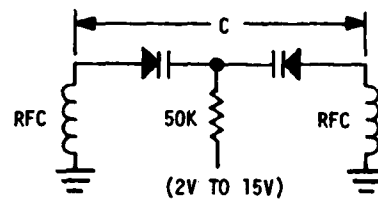
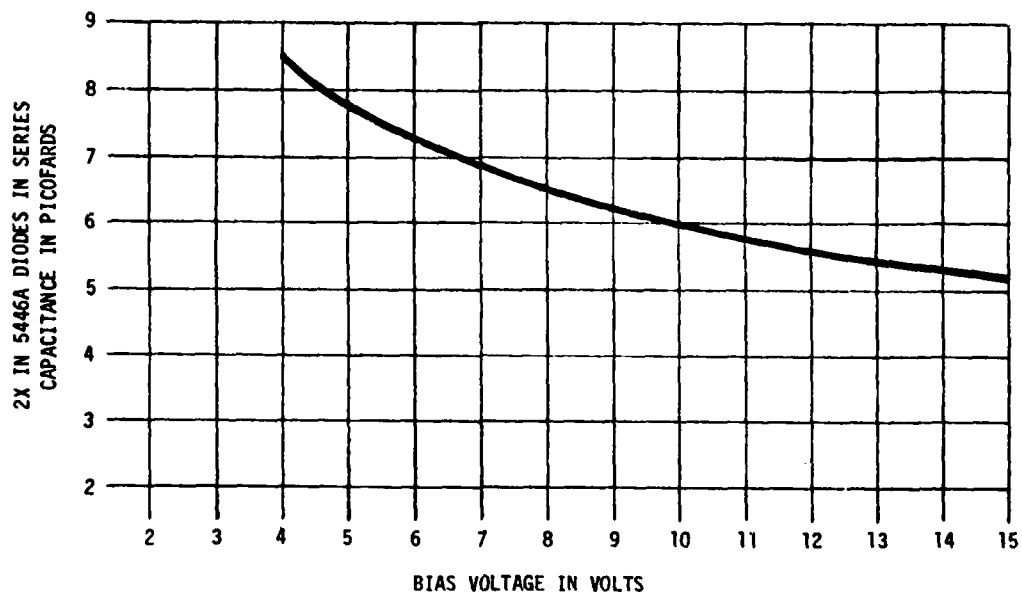


Figure 8. Capacitance of a Typical 1N5446A Varactor Diode Pair in Series vs Applied Voltage

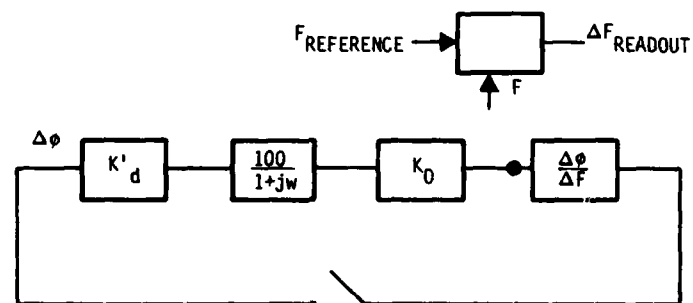


Figure 9. Incremental Open-Loop Model of Crystal Phase Lock-Loop

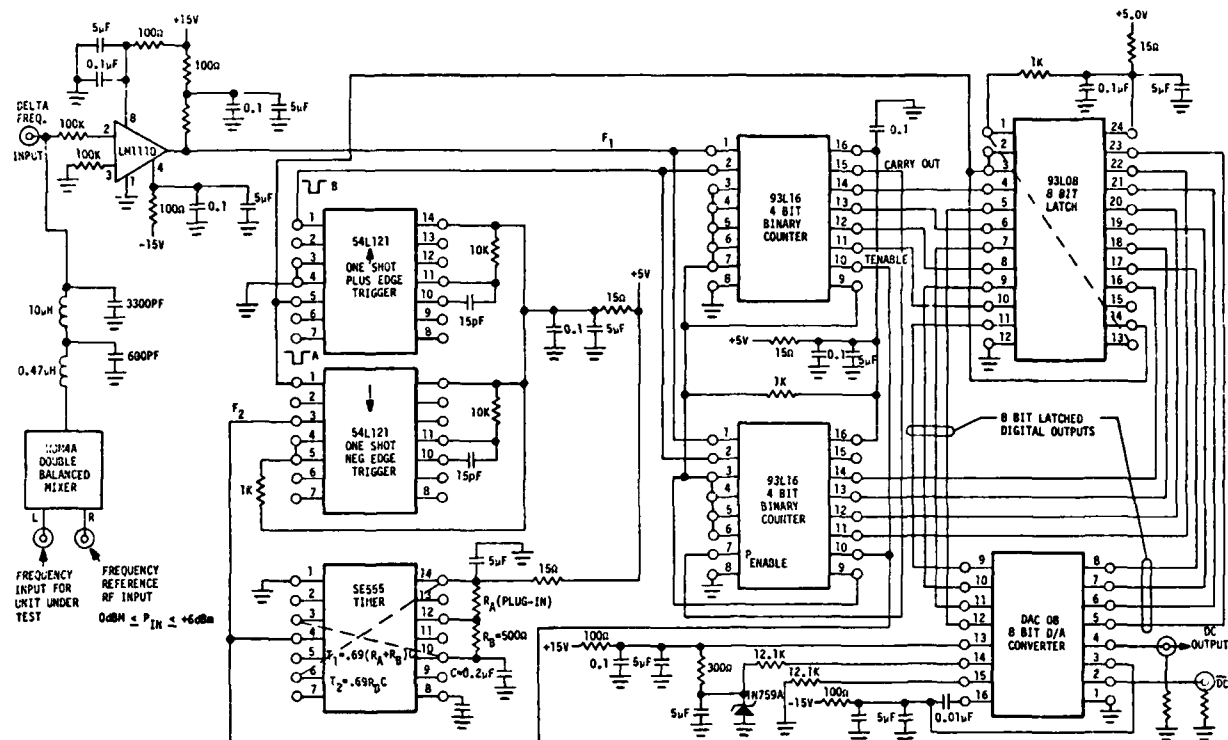


Figure 10. Linear Counter Type Frequency Discriminator

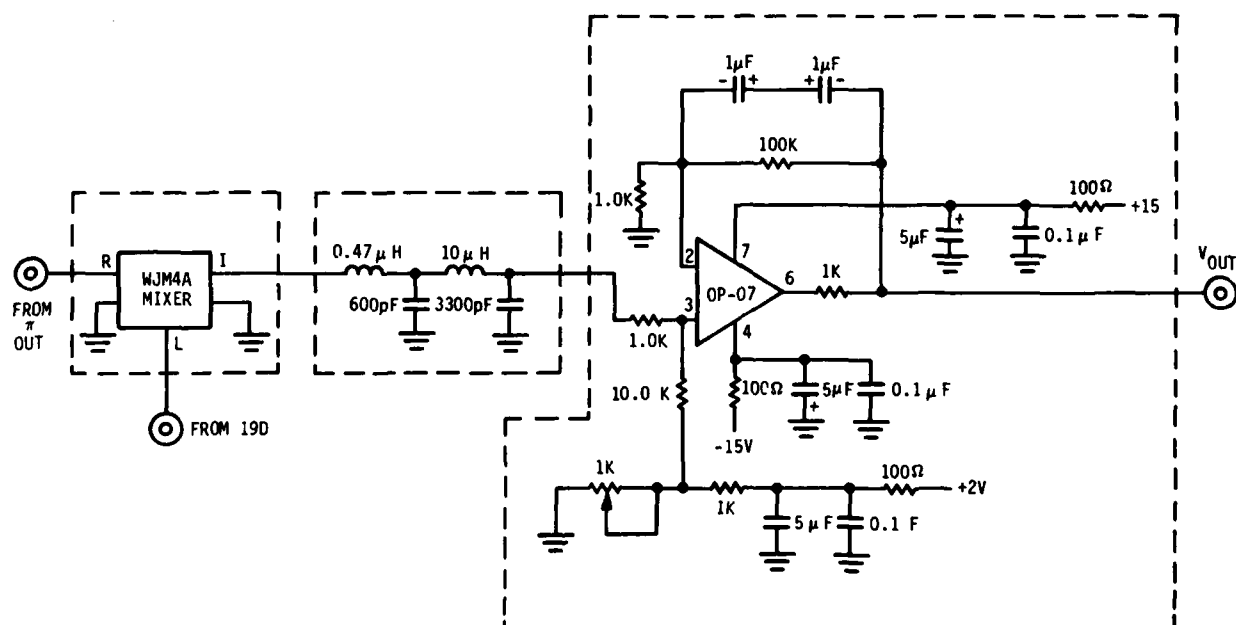


Figure 11. Coherent Detector

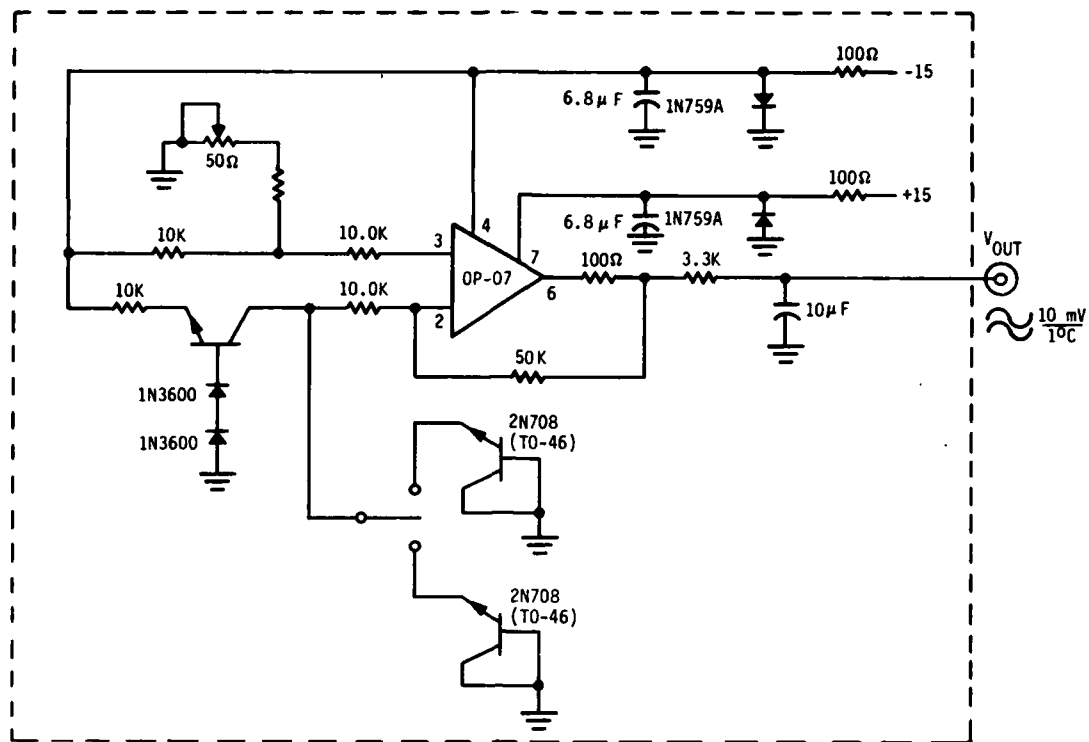


Figure 12. TO-46 Transistor Emitter-Base Diode Temperature Monitor
Adjusted to Provide a Readout in Degrees Centigrade

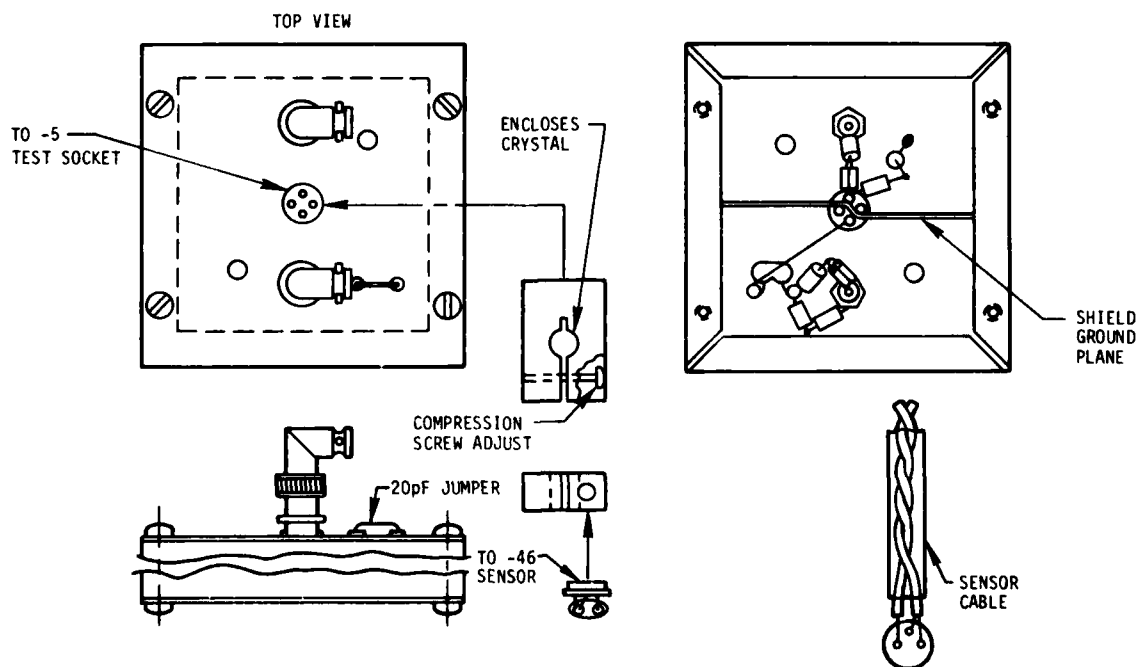
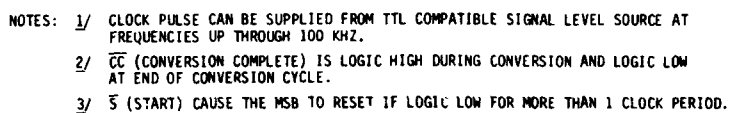


Figure 13. Illustrating a Method Used for Monitoring Crystal Temperature



NOMINAL FREQUENCY = 20835000 HZ .

UNIT NO. 102,3,4
PROJECT NO. 5823
DATE: 11/08/76

DELTA F (HZ)

TEMPERATURE DEGREE C

NOTE: 1/ ΔF $\left. \begin{array}{l} 20^{\circ} \\ 19^{\circ} \text{C} \end{array} \right\} = 3.6 \text{ PPM SERIAL NO. 102}$

2/ ΔF $\left. \begin{array}{l} 7^{\circ} \text{C} \\ 6^{\circ} \text{C} \end{array} \right\} = 9.6 \text{ PPM SERIAL NO. 104}$

POS. NO. 1,2,3

92

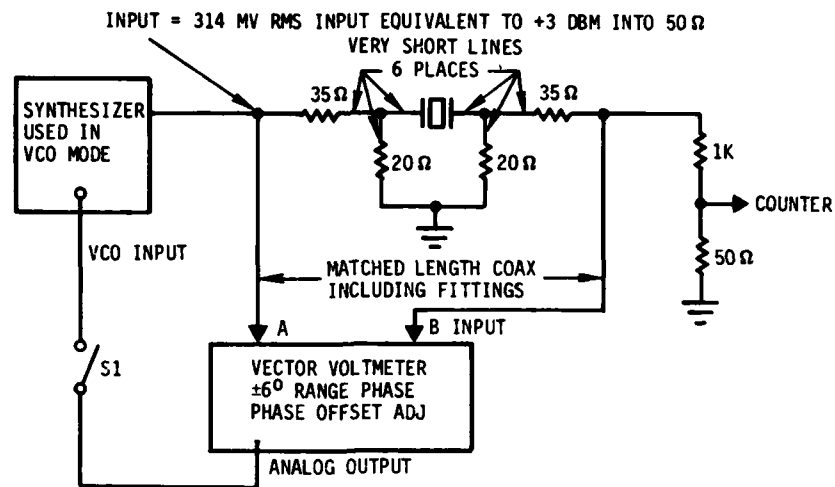


Figure 16. Basic Crystal Excitation Loop and Calibration Procedure

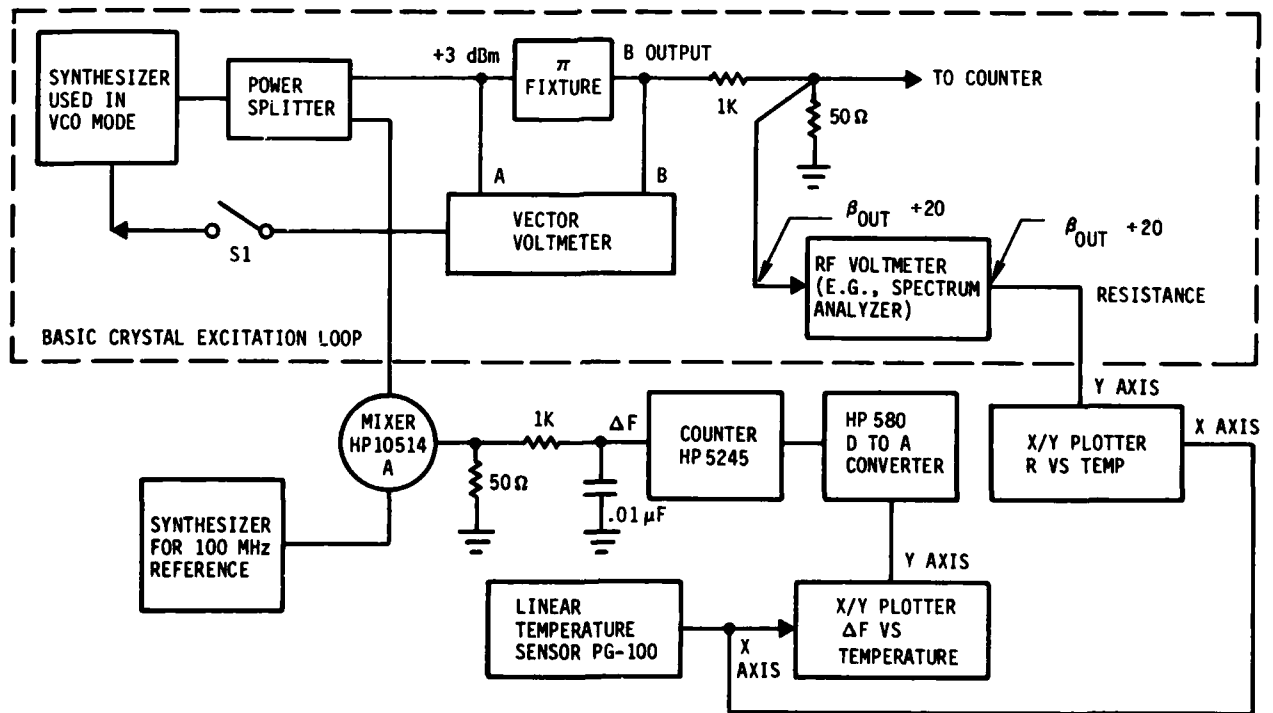
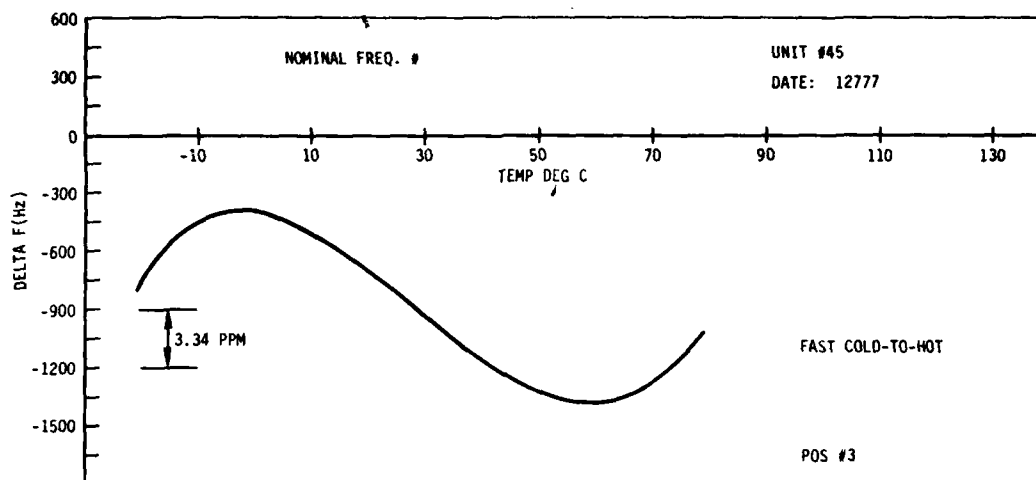
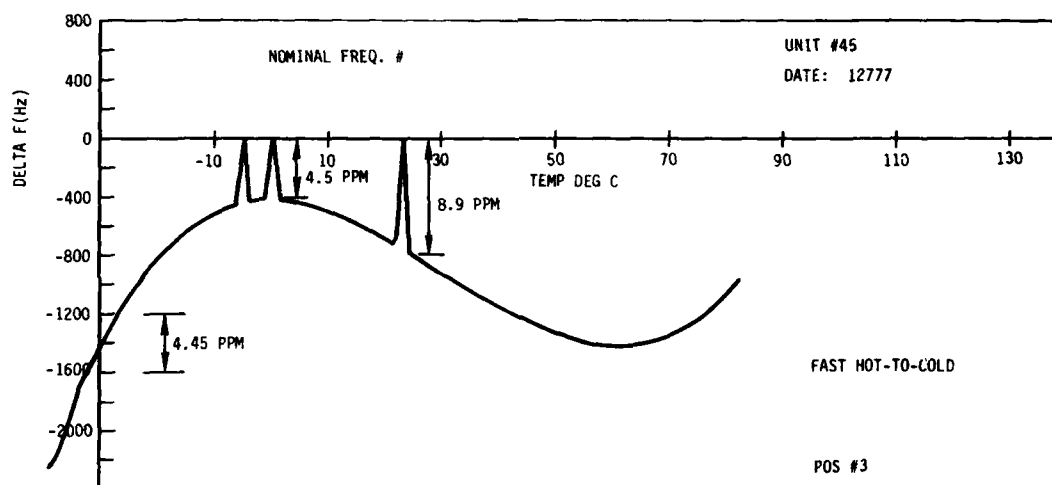


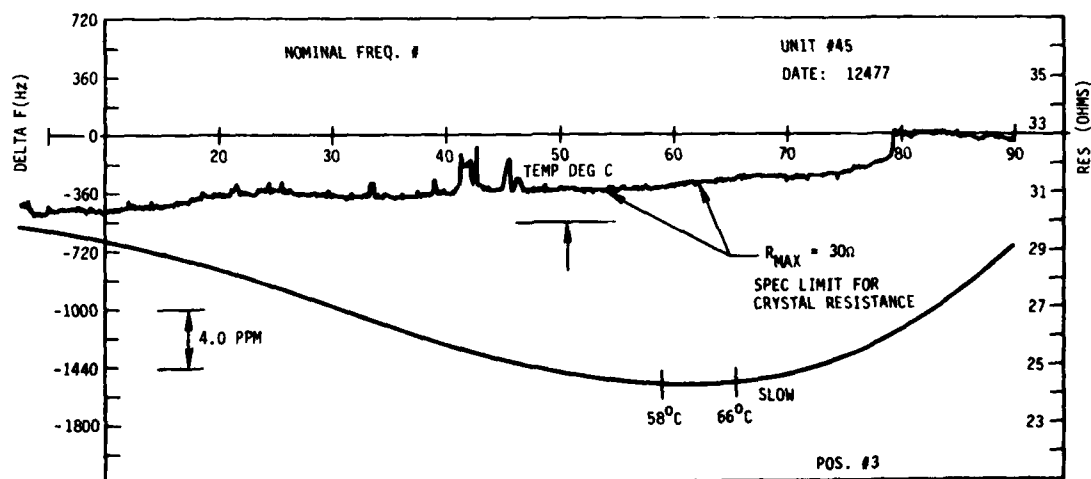
Figure 17. Complete Test Circuit - Frequency/Resistance vs Temperature for a 100 Megahertz Quartz Crystal



a)

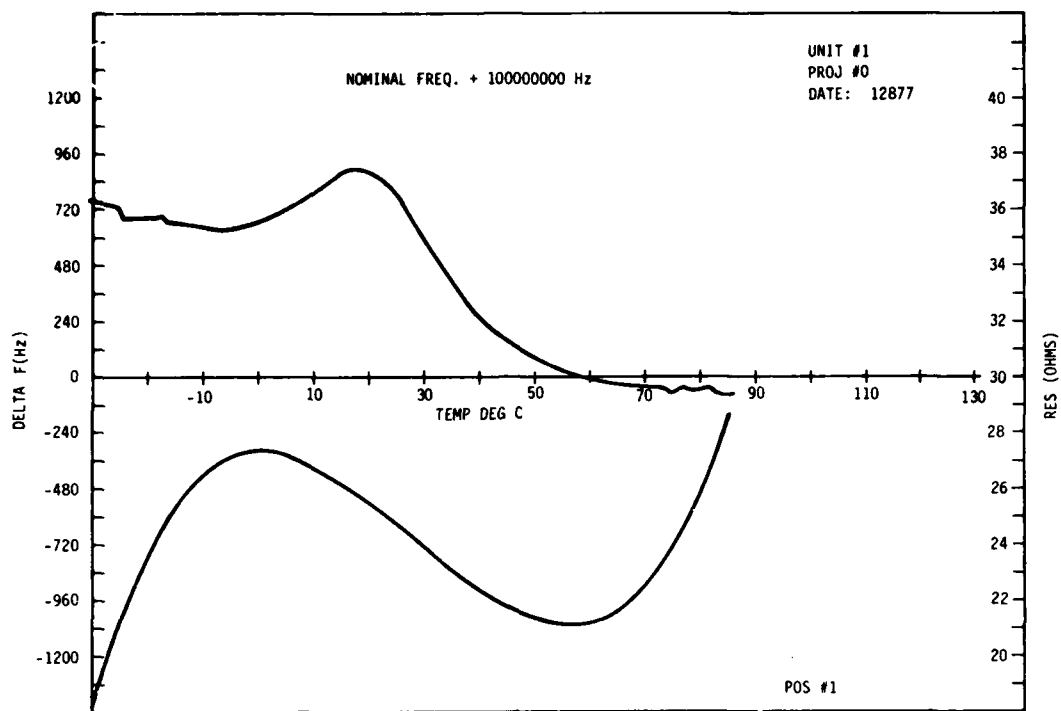


b)

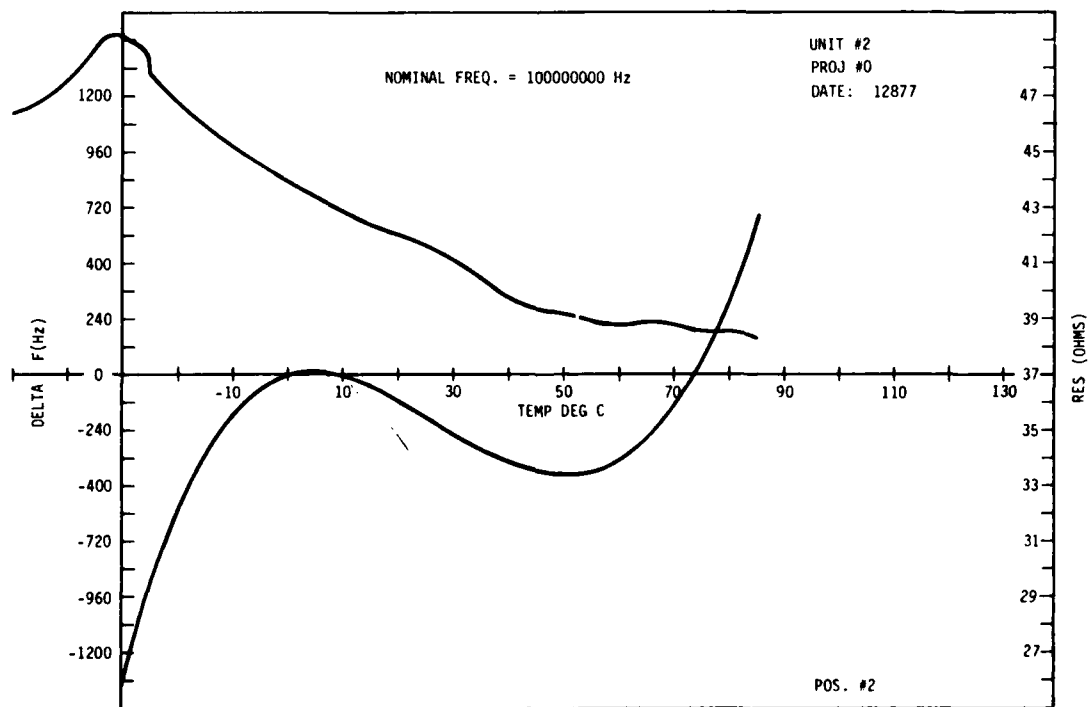


c)

Figure 18. Effects of Temperature Slew Rate on a Crystal Response



a)



b)

Figure 19. An Example of Crystal Resistance Change with Frequency
Where the Crystal Follows a Classic Cubic Response

EXTENDING THE FREQUENCY RANGE OF THE TRANSMISSION LINE METHOD FOR THE MEASUREMENT OF QUARTZ CRYSTALS UP TO 250 MHz

R. Fischer, L. Schulzke

KVG - Germany

Summary

The transmission line method according to IEC 444 has proven in practice as an efficient way to measure quartz crystal parameters in the frequency range from 1 to 125 MHz. Accuracy and repeatability of measurement are superior to any known active method with test oscillators and come close to the results obtained with sophisticated bridge arrangements, which are still the standard for fundamental laboratory work. The closed-loop transmission line system is, in conjunction with a programmable synthesizer and phase meter, perfectly adaptable to a computerized high-speed measuring system for production and quality control.

With increasing frequency the complex admittance resonance circle of the crystals 2-port equivalent circuit moves away from the real axis, as shown in Fig. (1); for frequencies, where the susceptance B_0 of the static capacitance exceeds $1/2 R_1$, there exists no intercept with the real axis and hence no zero-phase transmission any more. Below this critical frequency, f_r and R_r , per definition the characteristic low impedance figures for zero-phase transmission, can be measured easily, but determination of the motional parameters C_1 , L_1 and R_1 becomes strongly erroneous due to the influence of C_0 .

Compensation of C_0 by a tuned parallel inductance was proposed to overcome this problem. Provided the Q of this compensation inductance is inherently high, this C_0 -cancellation returns the centerpoint of the admittance circle exactly onto the real axis and direct access to the motional branch parameters of the crystal is given.

This paper describes an electrically tuned compensation circuit, which is directly connected to the knife-edge contacts of the π -network and permits cancellation of C_0 over a frequency range from 100 to 250 MHz. It consists basically of a symmetric varicap capacitor and a parallel-line inductance with a simultaneously tuned shortening capacitor at its end. Besides the circuit analysis a detailed description of the calibration and measuring procedure is given for the compensated transmission line network taking into account the finite Q of the compensation circuit and its influence on the determination of frequency and motional parameters.

Although this compensation method is not yet evaluated to its full extent, especially with respect to distributed reactances, it seems to be a promising solution for the measurement of VHF-crystals by the transmission line principle and can be easily adapted to existing equipment.

Introduction

The commonly used layout of a transmission line setup for the measurement of crystal parameters in the frequency range from 1 to 125 MHz is shown in Fig. (2): The RF-generator, mostly a synthesizer with a selectable decade stage running in VCO-mode, feeds via a power splitter the two channels A and B. Channel A is the reference path for the phase meter, whereas channel B keeps the transmission line network. Both channels are trimmed to the same electrical length and have 50 ohm terminations, where the test probes of the vector voltmeter are attached. The residual phase vs. frequency change in the order of 1 to 2 degrees over the full frequency range can be corrected by the phase offset control of the vector voltmeter. The servo output of the phase meter is fed to a search and lock circuit, which generates the VCO-control voltage of the closed-loop system for zero phase difference between the channels A and B.

The electrical configuration of the 2-section transmission line network is given in Fig. (3): The resistance values are designed for a maximal reflection coefficient of 2 % and power matching is obtained with a crystal resistance of $R_r = 25$ ohms. With reference to the graph in Fig. (3) variations of R_r to higher figures have a less pronounced effect onto the crystal drive level, although a higher resolution, i.e. phase vs. frequency change, is achievable with lower impedance networks.

Network calibration correctly should be done by insertion of a frequency compensated resistor, practical use however, with only slight degradation in accuracy, is insertion of a shortcut. After zeroing the phase meter the drive level is set to the estimated R_r of the crystal by setting V_{bs} to

$$V_{bs} = 0,183 (R_r + 25) \sqrt{\frac{P_c}{R_r}} \quad [1]$$

After inserting the crystal, its R_r can be calculated from the appropriate V_{bs} -reading by

$$R_r = \left(\frac{V_{bs}}{V_{bs0}} - 1 \right) \cdot 25 \quad [2]$$

and the actual power dissipation of the crystal becomes

$$P_c = 30,04 \cdot V_{bs}^2 \cdot \frac{R_r}{(R_r + 25)^2} \quad [3]$$

Until now anything is correct: By definition f_r and R_r are the characteristic frequency and resistance of the purely resistive crystal for the low-impedance case. Their deviation from the motional branch parameters f_s and R_s is systematic and can be taken into account. The normalized frequency offset between f_r and f_s is

$$\Omega = \frac{f_r}{f_s} - \frac{f_s}{f_r} = \frac{2r}{Q_0^2(1 + \sqrt{1 - 4R_1^2/X_0^2})} \quad [4]$$

$$r = C_0/C_1 \quad Q_0 = 1/\omega_s R_1 C_1 \quad X_0 = 1/\omega_s C_0$$

As long as the susceptance $B_0 = 1/X_0$ of the static capacitance is negligibly small, the motional branch parameters R_1 , C_1 and L_1 can be determined by the transmission line method. Under this provision is

$$f_r \approx f_s \quad R_r \approx R_1 \quad [5]$$

The effective Q of the crystal, loaded by the 25 ohms of the network then becomes

$$Q_{\text{eff}} = \frac{\omega_s L_1}{R_1 + 25} = \frac{1}{\omega_s C_1 (R_1 + 25)} = \frac{f_m}{\Delta f_{3\text{db}}} = \frac{f_m}{\Delta f_{\pm 45^\circ}} \quad [6]$$

Q_{eff} then can be calculated from the measurement of the relative bandwidth at 3 db attenuation or at $\pm 45^\circ$ phase offset. The motional capacitance and inductance then are

$$C_1 \approx \frac{\Delta f_{\pm 45^\circ}}{2\pi f_s^2 (R_1 + 25)} \quad L_1 \approx \frac{1}{\omega_s^2 C_1} \quad [7]$$

Especially for automated measurement processing a more accurate determination of the motional parameters can be derived from the analysis of the complex crystal impedance in series with its termination by the transmission line network. By taking 2 readings of V_b and frequency at opposite phase offsets of $\Delta\varphi$ one obtains

$$R_1 \approx \left(\frac{V_{b\pm\varphi} \cos \varphi}{V_{b(+\varphi)} + V_{b(-\varphi)}} - 1 \right) 25 \quad [8]$$

$$C_1 \approx \frac{(f_{(+\varphi)} - f_{(-\varphi)}) \cdot (V_{b(+\varphi)} + V_{b(-\varphi)})}{25\pi \sin \varphi V_{bs} (f_{(+\varphi)} + f_{(-\varphi)})^2}$$

$$L_1 \approx \frac{50 \sin \varphi V_{bs}}{\pi (f_{(+\varphi)} - f_{(-\varphi)}) \cdot (V_{b(+\varphi)} + V_{b(-\varphi)})}$$

Equations (5) through (8) are exactly valid only under the provision of a negligibly small susceptance $1/X_0$ or of C_0 being compensated. In other words: When the motional parameters are measured by opposite phase-offsets from zero-phase transmission, the resonance curve of the crystal becomes non-symmetric and one gets differing V_b -readings due to the vertical shift of the admittance circle. The theoretical errors resulting from this were computed by an analysis program of the complex crystal impedance loaded by the transmission line network, which was assumed to be purely resistive. The results are given

in Fig. (4) for crystals in the fundamental, 3rd and 5th harmonic mode. The errors become noticeable in the 3rd overtone and reach unacceptable figures in the 5th harmonic mode. These errors are principal ones and have nothing to do with the excellent correlation achieved between different transmission line setups.

Here we touch the core of the problem when discussing bridge vs. transmission line measurements: Not so much correlation of the frequency measurements is the problem, because of the high Q of the crystal the differences are sufficiently small for most practical purposes or can be corrected by an error analysis. Main problem is the erroneous determination of motional parameters by phase-offset methods due to the uncompensated C_0 , which must give physically wrong results in comparison to bridge measurements with increasing frequencies.

The bridge method eliminates this error by the initial balancing procedure and then measurement access is given directly to the motional arm. A way to solve this problem for the transmission line method was proposed by Franx (Ref. 1) by measuring the complex admittance by a multipoint method and then calculating the parameters, including C_0 . As a closed mathematical does not exist, the results have to be determined iteratively and the method is therefore restricted to systems linked to a high-speed computer. A further disadvantage is, that f_s and R_1 are calculated figures and cannot be measured directly.

Compensation of C_0 seems to be the simplest way to get direct access to the motional branch, the results obtained then should come closer to those of bridge measurements (Ref. 2), (Ref. 3).

Layout of the C_0 -Compensation Circuit

An attempt was made to construct a network with an adjustable C_0 -compensation. The evaluation was projected towards the measurement of VHF-crystals, where the conventional network, depending on its construction, has an upper frequency limit around 125 MHz.

Because the transmission line network is a symmetric device with reference from the crystal terminals to ground, the compensation circuit was designed for the same symmetry. When considering stray reactances of the crystal and the compensation circuit, these reactances are in a bridge arrangement and only their differences are of major importance.

The principle of the C_0 -cancellation is derived in Fig. (5): C_0 can be compensated according to Fig. (5.a) by a parallel inductance L_0 , which is resonant with C_0 at or close to f_s . Compensation is still achieved if L_0 is replaced by the parallel resonance circuit consisting of L_p and C_p (see 5.b), the resonance condition at f_s has then to be

fulfilled by $(C_p + C_0)$. Because of the above mentioned symmetry reasons the basic configuration according to Fig. (5.c) was chosen, where the electrical circuit is shown in Fig. (5.d) with varicaps instead of the variable capacitors C_a and C_b .

Actually the compensation circuit consists of the variable capacitance $C_a/2$ parallel to C_0 and the inductance $2L$, in series of which the simultaneously tuned shortening capacitor $C_b/2$ is placed. By folding the circuit about its plane of symmetry the equivalent schematic of Fig. (5.e) is derived; for the lossless case its complex impedance must be purely reactive and is

$$Z = j \cdot \frac{\frac{\omega L}{2} - \frac{1}{\omega^2 C_b}}{1 + \left(\frac{1}{2C_b} - \omega^2 \frac{L}{2}\right)(4C_0 + 2C_a)} \quad [9]$$

and the effective reactive components for the resonance condition at f_s can be written

$$L_p = \frac{L/2}{1 + \left(\frac{1}{2C_b} - \omega^2 \frac{L}{2}\right)(4C_0 + 2C_a)} \quad [10]$$

$$C_p = C_b \left[1 + \left(\frac{1}{2C_b} - \omega^2 \frac{L}{2}\right)(4C_0 + 2C_a) \right] \quad [11]$$

According to general network theory such a device has a pole impedance at zero frequency and a zero and pole with increasing frequency. By solving Eq. (9) for these conditions we obtain

$$\omega^2 = \omega_s^2 = 1/LC_b \quad [12]$$

for the series mode and

$$\omega^2 = \omega_p^2 = 1 / \left(LC_b \frac{2C_0 + C_a}{C_a + C_b + 2C_0} \right) \quad [13]$$

for the parallel resonance mode. Both modes are related to each other by

$$\omega_p = \omega_s \sqrt{1 + \frac{C_b}{C_a + 2C_0}} \quad [14]$$

$$\omega_p \approx \omega_s \sqrt{1 + \frac{1}{2}} = 1,22 \omega_s \text{ for } C_b \gg C_a \quad [15]$$

i.e. the unwanted series mode can be placed 20 % below the desired high-impedance parallel resonance mode, low enough not to affect the crystal measurement in the narrow bandwidth of consideration.

The mechanical layout of the compensation circuit is sketched in Fig. (6): The parallel-wire inductance is directly attached to the contact blades of the transmission line network. For the wire length l , the center-to-center distance a and the wire radius r (in millimeters) inductance of the parallel line for the same amounts of current flowing in

both wires can be calculated by

$$2L = 1 \left(4 \ln \frac{a}{r} \right) \cdot 10^{-9} \quad [16]$$

The tuning characteristic was analyzed by

$$C_v/pF = C_p + \frac{k}{(V_r + \Phi_k)^n} \quad [17]$$

C static capacitance (0.25 pF)
 V_r reverse voltage
 Φ_k contact potential (0.7 volts)

The planar type BB 109Q (Siemens) with $K=94.5$ and $n=0.90$ was chosen and the hyperabrupt MV 1403 (Motorola) with $K=920$ and $n=1.67$. For reverse voltages from 12 to 2 volts C_a changes from 19 to 77 pF and C_b from 18 to 175 pF. The measured characteristic in Fig. (7) shows, that for C_0 up to 7 pF the range from 100 to 250 MHz is covered.

By defined capacitive detuning and observation of the resonance shift the effective reactances were measured

$$C_p = \frac{\Delta C}{\left(\frac{1}{1 - \Delta f/f_0} \right)^2 - 1} \quad L_p = \frac{1}{4\pi^2 f_0^2 \cdot C_v} \quad [18]$$

and plotted in Fig. (8): The chosen configuration undergoes a 1:5 change of inductance L_p and has a relatively stable S-shaped effective capacitance C_p .

The loaded Q_p of the compensation circuit can be measured by plotting the phase vs. frequency characteristic, an example is given in Fig. (9) for 166 MHz: The maximum slope at resonance is $4,1 \cdot 10^{-5}$ degrees/Hz, if we assume a crystal with $Q = 50.000$ the measurement over its bandwidth would create a total phase error of 0,14 degrees originating from the off-resonance reactances of the compensation circuit. As Q is directly related to phase change, the low Q_p of the compensation circuit affects phase measurements of the crystal only by a factor Q_p/Q_c .

Under these assumptions the compensation circuit can be replaced by its parallel loss resistance R_p , and this can be considered to be frequency independent within the frequency range of consideration. With this statement measurement and calculation of the motional parameters is greatly simplified. In Fig. (10) Q_p and R_p are plotted over the frequency range considered here: R_p figures in excess of 1000 ohms are achieved with a maximum of Q_p at 225 MHz, which of course would be better placed in the middle of the operating frequency range.

Calibration and Measuring Procedure

Because of the above reasons the low- Q compensation circuit is replaced in the following considerations by its parallel loss resistance R_p and this is assumed to be purely resistive and frequency independent. Under

this provision the crystal with the parallel compensation circuit and the terminating resistance R_t of the transmission line network can be comprehended to the equivalent circuit of Fig. (11): The complex impedance Z_m of the motional arm of the crystal can be written in the form

$$Z_m = R_1 + j \omega_s L_1 \Omega \quad [19]$$

$$\text{where } \Omega = \frac{\omega}{\omega_s} - \frac{\omega_s}{\omega} \quad \omega_s^2 = 1/L_1 C_1$$

introducing $\omega = \omega_s + \Delta\omega$ $\Delta\omega/\omega_s \ll 1$

Z_m of Eq. (19) reduces to

$$Z_m = R_1 + j 2 \Delta\omega L_1 \quad [20]$$

The complex impedance of the compensated crystal in series with the terminating resistance R_t of the network now becomes

$$Z = R_t + \frac{R_p [(R_t^2 + R_1 R_p + 4 \Delta\omega^2 L_1^2) + j (2 \Delta\omega L_1 R_p)]}{(R_1 + R_p)^2 + (2 \Delta\omega L_1)^2} \quad [21]$$

The phase angle of this impedance is

$$\varphi = \arctan \frac{(2 \Delta\omega L_1)^2 \cdot (R_p + R_t) + R_p R_t^2 + R_p^2 R_1 + R_t (R_1 + R_p)^2}{(2 \Delta\omega L_1) \cdot R_p^2} \quad [22]$$

and its maximum

$$\varphi_{\max} = \arctan \frac{\sqrt{4B(R_p + R_t)}}{R_p^2} \quad [23]$$

$$B = R_p R_1^2 + R_1^2 R_p + R_1^2 R_t + 2 R_1 R_t R_p + R_t R_p^2 \quad [24]$$

From Eq. (22) now the motional inductance of the crystal can be extracted

$$L_1 = \frac{R_p^2 \operatorname{ctg} \varphi \pm \sqrt{R_p^4 \operatorname{ctg}^2 \varphi - 4B(R_p + R_t)}}{4 \Delta\omega (R_p + R_t)} \quad [25]$$

For the low impedance case around series resonance the positive sign of the root is related to negative phase angles and vice versa.

The calibration and measuring procedure with the compensated transmission line network is carried out in the following steps:

Step 1: With a shortcut inserted into the network the drive level is set by adjusting V_{bs} according to Eq. (1).

Step 2: After inserting the crystal the generator is set slightly below the crystal frequency (about 100 kHz). C_0 is then cancelled out by tuning the compensation

circuit to minimum V_b or zero phase reading. The parallel loss resistance R_p is then calculated from

$$R_p = \left(\frac{V_{bs}}{V_{b(\min)}} - 1 \right) R_t \quad [26]$$

Step 3: Next the generator is tuned to the crystal frequency and locked at zero phase, then is

$$f \Rightarrow f_s \quad R_1 = \frac{Z_0}{1 - Z_0/R_p}$$

$$Z_0 = \left(\frac{V_{bs}}{V_{b(\varphi=0)}} - 1 \right) R_t \quad [27]$$

where Z_0 is the resistance resulting from paralleling R_1 and R_p in the zero phase condition.

Step 4: Finally two frequency measurements are made at identical positive and negative phase offsets, the offset angle preferably should be not larger than half of the figure maximally attainable. The remaining motional arm parameters are then calculated from

$$L_1 = \frac{R_p^2 \operatorname{ctg} \varphi - \sqrt{R_p^4 \operatorname{ctg}^2 \varphi - 4B(R_p + R_t)}}{4\pi \cdot (f_{(+\varphi)} - f_{(-\varphi)}) (R_p + R_t)}$$

$$C_1 = 1/\omega_s^2 L_1 \quad Q = \omega_s L_1 / R_1 \quad [28]$$

with B taken from Eq. (24).

Although these formulas are looking somewhat lengthy for practical use, the computations are a matter of seconds with the aid of a programmable pocket calculator.

Conclusion

This paper shall be an attempt to improve the principle of the transmission line method for the measurement of quartz crystals in 3 respects:

1. To extend the frequency range for the measurement of VHF-crystals beyond the limit given by the susceptance of C_0 .
2. To evaluate a system which gives results close to the actual application, as C_0 -compensation is widely used in VHF-crystal oscillators.
3. To overcome the problem of erroneous measurement of the motional parameters by the widely used phase offset method.

The third point until now was considered for VHF-crystals only. With the results obtained we now think to apply this principle also at frequencies below 100 MHz to achieve a better agreement with bridge measurements.

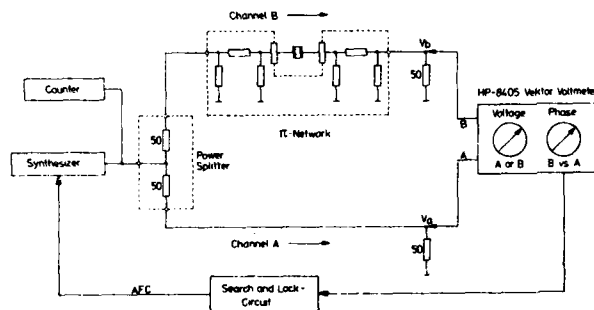


Fig 2: Commonly used Transmission-Line Network Setup.

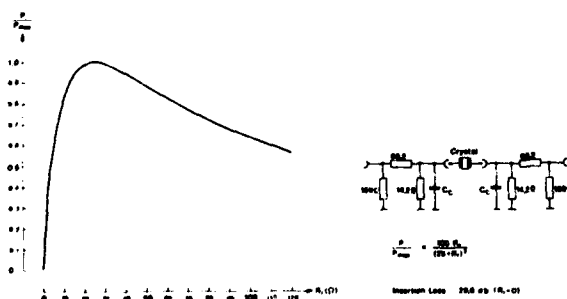


Fig.3: TL-Network Resistance Figures and Crystal Power Dissipation vs R_r Characteristic.

References

- 1: Report on Basic Methods of Measurement of the Parameters of the Equivalent Circuit of Quartz Crystals and Piezoelectric Resonators. Private Communication (1971).
- 2: E.Hafner, A.Ballato, P.Blomster: Quartz Crystal Measurements. Proc. of the ASFC (1970), pp. 177-190
- 3: E.Hafner, W.J.Riley: Implementation of Bridge Measurement Techniques for Quartz Crystal Parameters. Proc. of the ASFC (1976), pp. 92-102

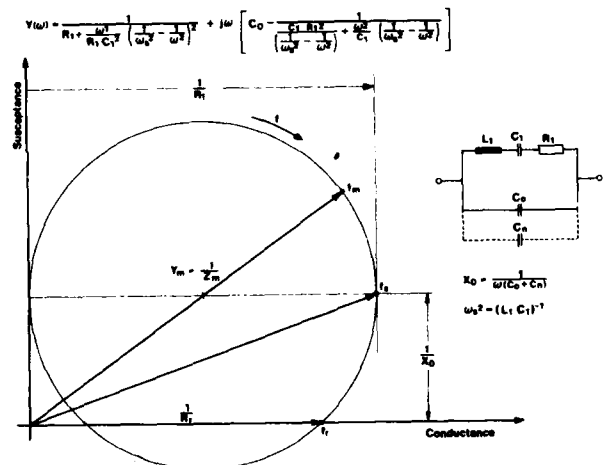


Fig.1: Complex Admittance Resonance Circle Diagram.

	Fundamental			3rd Overtone			5th Overtone		
f [MHz]	10	20	30	30	60	90	90	180	180
Q [10 ³]	50	50	50	127	63	42	92	46	31
Q·f [10 ¹⁰]	0.5	1.0	1.5	3.8	3.8	3.8	4.6	4.6	4.6
C0 [pF]	6	6	6	6	6	6	6	6	6
C1 [1F]	27	27	27	2.4	2.4	2.4	0.65	0.65	0.65
R1 [Ω]	11.8	5.90	3.93	17.5	17.5	17.5	53.2	53.2	53.2
$\frac{L_1 - R_1^2}{C_1}$ [ppm]	0.044	0.044	0.044	0.079	0.313	0.702	0.539	2.098	4.524
$\frac{R_1 - R_1^2}{C_1}$	$-1.7 \cdot 10^{-8}$	$-1.9 \cdot 10^{-8}$	$-2.0 \cdot 10^{-8}$	$-4.0 \cdot 10^{-8}$	$-1.6 \cdot 10^{-8}$	$-3.5 \cdot 10^{-8}$	$-9.9 \cdot 10^{-9}$	$-3.7 \cdot 10^{-9}$	$-2.7 \cdot 10^{-9}$
$\frac{C_1 - C_1}{C_1}$ (Note 1)	$< 10^{-8}$	$2.9 \cdot 10^{-8}$	$7.0 \cdot 10^{-8}$	$6.3 \cdot 10^{-8}$	$2.5 \cdot 10^{-8}$	$5.7 \cdot 10^{-8}$	$7.5 \cdot 10^{-8}$	$2.9 \cdot 10^{-8}$	$6.2 \cdot 10^{-8}$
$\frac{C_1 - C_1}{C_1}$ (Note 2)	$< 10^{-8}$	$3.8 \cdot 10^{-8}$	$2.5 \cdot 10^{-8}$	$-2.5 \cdot 10^{-8}$	$-1.4 \cdot 10^{-8}$	$-3.1 \cdot 10^{-8}$	$-8.1 \cdot 10^{-8}$	$-3.2 \cdot 10^{-8}$	$-2.2 \cdot 10^{-8}$

①: C_1' calculated from Eq. (7)

②: C_1' calculated from Eq. (8)

Fig.4: Systematic Error of Measurements with uncompensated TL - Network (calculated).

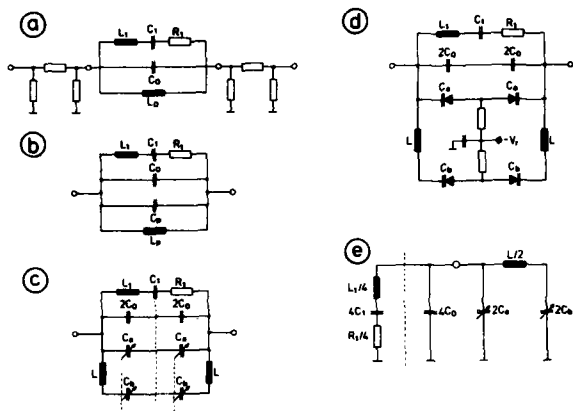


Fig. 5: Principle of C_0 -Compensation attached to the TL-Network

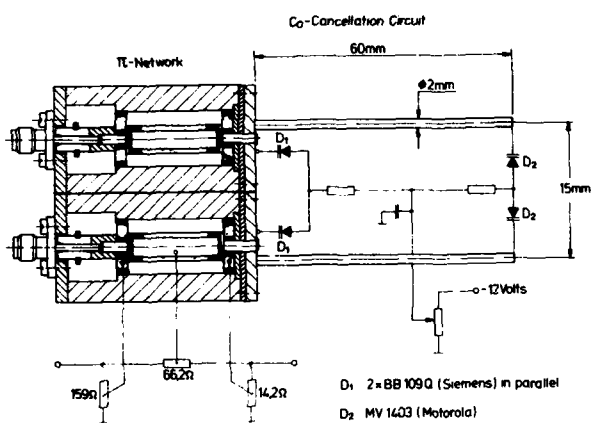


Fig. 6: Simplified Schematic of the TL-Network with C_0 -Cancellation Circuit.

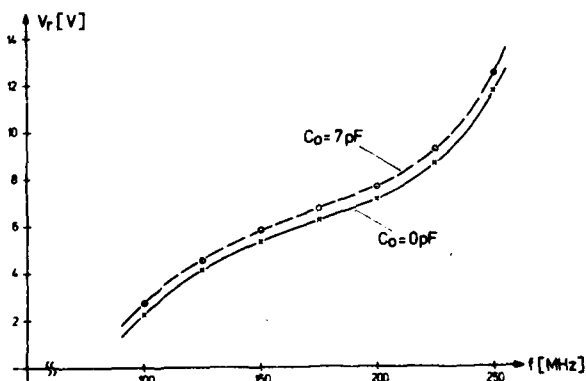


Fig. 7: Tuning Characteristic of the C_0 -Compensation Circuit.

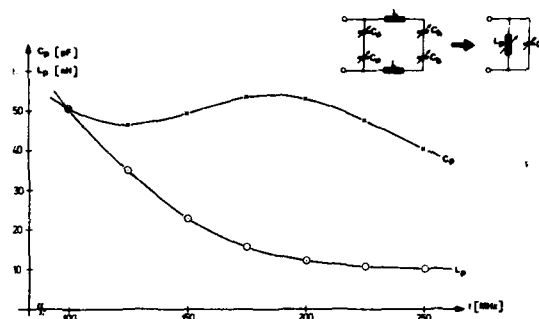


Fig. 8: Effective Inductance and Capacitance of the Compensation Circuit

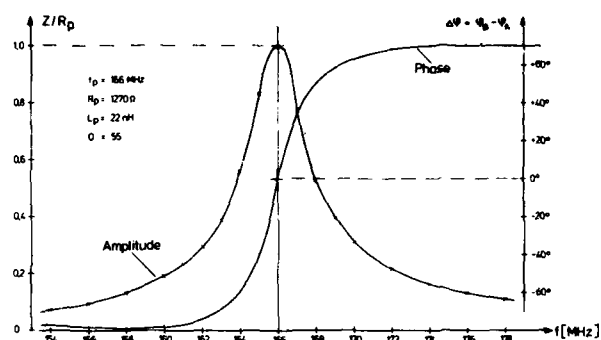


Fig. 9: Phase and Amplitude Transmission of the compensated TL-Network

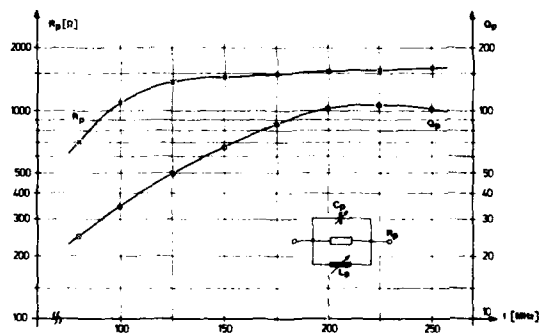


Fig. 10: Q_p and Parallel Loss Resistance R_p as a Funktion of Frequency

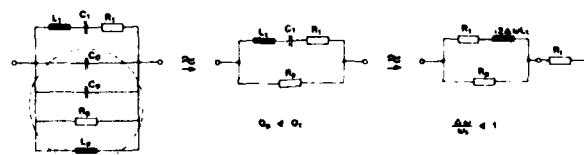


Fig. 11: Simplified equivalent Circuit of C_0 -compensated Crystal and TL-Network

OVENLESS ACTIVITY DIP TESTER

A. Ballato and R. Tilton

US Army Electronics Technology and Devices Laboratory (ECOM)
Fort Monmouth, New Jersey 07703

Summary

Anomalies in the frequency- or admittance-temperature characteristic of a resonator are called "activity dips" or "bandbreaks." These anomalies are generally conceded to be caused by various combinations of different modal frequencies coming into coincidence at particular temperatures because of differing temperature coefficients. The presence of activity dips is a persistent problem and necessitates a good deal of costly testing for medium- and high-precision resonator units.

The modal interference that takes place may be either linear or nonlinear, and a variety of mechanisms leading to this phenomenon have been reported.¹⁻⁶

The effect of inserting a capacitor in series with a crystal is to shift all resonance frequencies upward by amounts roughly inversely proportional to the capacitance ratios of the modes involved. As a result of this and because the interfering modes have differing temperature coefficients and capacitance ratios, it is possible to alter the temperature at which a dip is produced. By using an adjustable capacitance network, the activity dip spectrum may be electronically swept while the temperature remains constant. The time-consuming and costly frequency-temperature oven runs presently used for detecting activity dips may thereby be replaced by an all-electronic, constant-temperature procedure that is rapid, simple, and well adapted to microprocessor control.

This paper outlines the theory of operation, the required measurement setup, and sample results on crystal test units.

Introduction

Anomalies in the frequency- or admittance-temperature characteristic of a resonator are called "activity dips" or "bandbreaks." At least two distinct species exist: "design-related" dips and "process-related" dips. The latter type arise from shortcomings in the processing phases, where, for example, an improperly deposited electrode film peels or blisters in a reversible, temperature-dependent manner; these are not considered further here. "Design-related" activity dips are those for which the structural configuration remains unaffected by temperature changes, but instead depend solely upon geometry and material constants. The presence of "design-related" activity dips is a persistent problem and necessitates a good deal of costly testing for medium- and high-precision resonator units. Doubly rotated cuts generally could be expected to have even more problems in this regard than AT-cuts since they exhibit less symmetry and have, therefore, a more complicated mode spectrum when lateral boundaries are taken into account. An exception appears to be the SC/TTC orientation⁷⁻¹⁰ for which activity dips have yet to be encountered.^{4,11}

The modal interference that takes place may be linear or nonlinear. If the impressed voltage can drive the desired thickness mode and at the same time drive a harmonic of a flexural mode, e.g., then the vibrator admittance will reflect this fact as the

linear superposition of the separate modal admittances. With temperature changes it is possible for the two resonance frequencies to cross and produce an anomaly. Linear activity dips have been described by Wood and Seed⁴ and by Fukuyo *et al.*²

Nonlinear activity dips are less well understood and perhaps more important. Wood and Seed⁴ found the AT-cut fundamental thickness shear frequency to be affected by interfering modes at twice its frequency; Franx³ observed the same type of coupling due to a mode at three times the fundamental. Birch and Weston⁶ investigated both cases. Koga⁵ found the twenty-first harmonic of contour extension interfering nonlinearly with the thickness shear fundamental. Similar results were obtained by Fukuyo *et al.*² In all cases the sensitivity of mode coupling to power levels is a characteristic of the nonlinearity. Hafner¹ found that the anomalies encountered at the fifth and seventh harmonics were nonlinear in nature and depended on the electrode film as well as the quartz.

The effect of inserting a capacitor in series with a crystal is to shift all resonance frequencies upward by amounts roughly inversely proportional to the capacitance ratios of the modes. The temperature coefficients of the various modes are similarly altered. As a result of these facts, and because the interfering modes have differing temperature coefficients, it is possible to alter the activity dip temperature by a series load capacitance. By using an adjustable capacitance network, the activity dip spectrum may be electronically swept while the temperature remains constant. In effect, the time-consuming, costly, frequency-temperature oven runs presently used for detecting activity dips are replaced by an all-electronic constant-temperature procedure that is rapid, simple, and well adapted to microprocessor control.

Frequency-Temperature Spectrum

In Figure 1 we show a portion of the unwanted mode spectrum of a quartz resonator, as a function of temperature, obtained by Fukuyo *et al.*² This painstaking work shows clearly how various responses may drift into coincidence at particular temperatures due to differing temperature coefficients of frequency. The responses shown are those detected piezoelectrically. It is also possible to drive indirectly nonpiezoelectric modes that would not appear on such a chart and yet whose presences could produce anomalies. This is particularly true of nonlinearly coupled modes.

Series Load Capacitance

In the vicinity of the main, desired mode, a crystal resonator is adequately represented by the Butterworth-Van Dyke equivalent circuit in Figure 2. The effect of inserting a load capacitor C_L in series with the crystal may also be represented by a circuit of the same form, but with modified values. The appropriate formulas are also given in Figure 2.

The important fact is that the crystal resonance frequency $f_R = f_g$ is shifted by the load capacitor to a higher frequency f_L . The load frequency f_L depends

upon C_L through the fraction α and upon the crystal capacitance ratio r . For an AT-cut operated at the fundamental, a typical value for r is 250, so that f_L could be pulled up to 2,000 ppm by extreme changes in C_L . An interfering mode, on the other hand, is apt to have a very large capacitance ratio and so would be shifted very little by the presence of C_L . This would be the case with modes such as the high overtones of flexure¹²; nonpiezoelectric modes would be completely unaffected by the presence of C_L . Because the effective capacitance of the thickness modes, such as the desired AT-cut shear mode, varies with the square of the harmonic, the overtones will shift correspondingly less than the fundamental. The same is true of their nonlinearly produced subharmonics.

The series load capacitance thus shifts upward the different modes by different amounts. Accordingly, the points of intersection of the modes (cf. Figure 1) will also change. This is the basis upon which the test method rests.

Figure 3 contains "hot plots" made of a 20 MHz resonator. The upper S-shaped curves are plots of frequency versus temperature and are typical of AT-cuts. The lowest S-curve was measured without C_L and labeled f_R . It has an anomaly marked by an arrow at about 78°C. The corresponding curve of grid current (proportional to the crystal admittance) is marked R_R , and also exhibits an anomaly. When a load capacitor C_{L1} is placed in series with the crystal in the oscillator, the curves f_{L1} and R_{L1} are produced. Now the activity dip has been displaced downward in temperature to approximately 58°C. Substitution of a smaller capacitor C_{L2} is seen to shift the activity dip anomaly further downward in temperature to about 44°C. The downward shift in temperature for the anomaly means that the temperature coefficient of the interfering mode is negative. Assuming that the resonator in question had $r = 250$ would mean that the interfering mode had a temperature coefficient of about $-5.6 \times 10^{-6}/K$.

The $\Delta f/f$ scale provided on Figure 3 pertains to each of the S-curves individually; the vertical separation between S-curves has been greatly reduced so that they could be compared more easily. The resistance curves have been similarly displaced for clarity, and relative position is unimportant.

In order to see more clearly how the activity dip temperature shifts refer to Figure 4; at the bottom is a frequency-temperature curve of a crystal vibrator at resonance, without C_L . On the scale of the drawing the curve is nearly horizontal; so to keep in mind its S-shaped familiar character, its vertical scale has been exaggerated by a factor of ten. Two curves representing unwanted, interfering modes have been drawn to intersect the S-curve at a common point located at 80°C. With a series capacitor attached, the S-curve moves upward by 2000 ppm and the points of intersection now occur at lower temperatures. The mode with coefficient of -60 ppm/K produces a dip at 48°C, and the -20 ppm/K mode would produce a dip instead at -23°C .

It is important to note that the interfering mode need not be actually present as shown in Figure 1 or Figure 4. It is only necessary that the harmonic or subharmonic producing the disturbance be represented as shown in Figure 4; the curve would thus be a "virtual response" on the figure. For example, Franx³ observed nonlinear coupling due to a mode at three times the fundamental. If this interfering mode were to have its temperature behavior mapped and then have its frequency divided by a factor of three, it could then be drawn as a "virtual response" on a figure

along with the fundamental curve for the purpose of determining how the dip temperature would change with C_L . The "virtual" curve would appear like one of the straight lines in Figure 4, whereas the actual interfering mode would appear at some harmonic frequency off the scale of the plot.

Activity Dip Scanner

Our observations concerning the effects of putting a load capacitor in series with a crystal resonator in an oscillator provide a method for trading off temperature scans, in search of activity dips, against load capacitor scans. That is, adjustment of C_L can be used to replace changes in temperature.

A series C_L is currently made use of for:

- Measuring crystal C_L ;
- FM, by using a varactor;
- Compensating the frequency-temperature crystal characteristic in TCXO's.

We now propose to use the effect to permit electronic scanning of activity anomalies. A block diagram of an activity dip scanner is shown in Figure 5.

The program control+signal processing (PCSP) block is implemented either by a microprocessor or by a programmable calculator and data interface. It controls the sweep generator as well as receives resistance and frequency information for processing. A completely instrumented system would include provisions for acceptance/rejection of the unit for activity dips, evaluation of crystal parameters, output of results, and control of the crystal unit input and output flow through the scanning apparatus.

The sweep generator supplies the sawtooth voltage to both the X-axis of the scope and the varactor network. The varactor network responds to the voltage changes of the sweep generator and produces the variable C_L which is in series with the crystal unit under test.

The oscillator requires care in its design, but details will not be gone into here. One must be sure that the sought-for anomalies originate in the crystal and not in the oscillator, so that unwanted frequency components must be strongly discriminated against. Depending on the type of crystal units and the stringency of the testing required, it may be adequate to use a standard CI Meter as the oscillator. We need only assume that the oscillator is such that it adjusts its frequency such that the crystal-load capacitor combination operates near or at its zero reactance point at f_L . Provisions for maintaining the drive level of the crystal constant as C_L is varied should also be considered in any embodiment of Figure 5.

The oscillator frequency output is mixed with a reference oscillator operating at the nominal frequency of the crystal under test. The difference frequency is applied to a frequency discriminator, whose output is displayed on the Y-axis of the scope.

A system simpler than that of Figure 5 results from the elimination of the PCSP unit, and operation of the sweep generator in a free-running mode.

The following instruments were used in a simple test setup for determining the feasibility of the method:

- Sweep generator: HP 3300A Function Generator
- rf oscillator: Crystal Impedance Meter TS-683/TSM (with low drive modification)
- Frequency discriminator: GR 1142-A Frequency Meter and Discriminator
- Reference oscillator: HP 5105A Frequency Synthesizer and HP 5110B Synthesizer Driver
- Mixer: HP 10514A Double Balanced Mixer.

Examples of the type of outputs obtained with the simplified setup are shown in Figures 6, 7, and 8. These show line drawings taken from actual scope photos. In all cases the temperatures T_1 and power levels P_1 follow the relations: $T_1 < T_2 < T_3$ and $P_1 < P_2 < P_3$. In Figure 6 temperature is held constant and power level is varied. At the low level P_1 the activity dip shows up as a sudden change in slope, although the curve remains a single-valued function. With increased drive, the dip is manifested by the presence of a hysteresis loop. Arrows indicate the direction of traverse as varactor voltage is swept back and forth.

When power level is held constant and temperature allowed to vary, curves such as those in Figure 7 result. Here the hysteresis loop position changes, moving down the curve with increasing temperature. This indicates that the interfering mode has a negative temperature coefficient of frequency. One may see this as follows: C_L decreases with increasing varactor voltage, and a decrease in C_L increases the shift of f_L from f_R . Therefore, high varactor voltages mean large f_L shifts. If the anomaly occurs at large varactor voltage at T_1 and again at a lower voltage at a higher temperature T_3 , then (cf. Figure 4) the coefficient of the undesired interfering mode is negative.

Figure 8 is comprised of a montage of graphs taken from measurement photos. Power is constant along rows; temperature is constant along columns. For the operation of the measurement setup it is not necessary to operate at power levels sufficient to produce a hysteresis loop. When no activity dip is present, the Δf - ΔV curve will be smoothly monotonic (it is not necessary that it be linear). With a dip present, the slope of the smooth curve is disturbed and indicates the existence of the anomaly. A simple differentiating network can be used to accentuate the "glitch," and detect the activity anomaly more readily.

Alternatively, the system can be simply arranged to use a voltage proportional to the crystal resistance ("grid current" in Figure 3) as the Y-axis scope input. In different cases the anomaly will be more prominent in either the frequency or resistance shift, and either may be used in the method.

Range Extension

A load capacitance by itself can only shift the oscillator frequency between the limits $f_0 \leq f < f_A$, that is, between resonance and antiresonance of the crystal unit. If, at the temperature of the test, the unwanted interfering mode causing the activity dip does not occur between these limits, then the anomaly will go undetected by the test.

Fortunately, there are simple ways of extending the range of the method. Figure 9 depicts qualitatively the effects on the pole-zero spectra resulting

from the placement of inductances in series and in parallel with the crystal. The series inductance ("stretching coil") lowers the first reactance zero. A parallel inductance shifts the pole at f_A upward. A series varactor with series/parallel inductor moves f_L from below f_R to above f_A .

The series and parallel inductances can be electronically switched in and out of the circuit using PIN diodes controlled by the PCSP unit in the automatic measurement mode.

From the point of view of network theory, the pole-zero shifts allowed are large; the limits imposed by the physics of the crystal vibrator are more stringent. Leaving aside all considerations of sensitivity of the system (Figure 5) on the detectability of activity dips, the major factors are the amplitudes of vibrator displacement and strain and the magnitude and type of the coupling mechanism to the interfering mode. These are discussed briefly in the appendices.

For a given resonator design, experience will dictate practical ranges for the test system parameters, based on the type of modal interferences encountered, design power level of crystal, etc.

Conclusion

A simple, rapid, electronic means of detecting the frequency/resistance anomalies known as activity dips has been described, with experimental confirmation. The method appears to be feasible to use in a number of possible ways:

- As a rapid go--no go 100% inspection method for eliminating units having obvious and easily detected activity dips prior to making costly temperature runs.

- As a more highly instrumented method combined with measurements for determining the crystal C_0 , C_1 , R_1 , L_1 parameters.

- As a hybrid method, combined with oven runs for interpolating between temperatures.

In this last variation, an oven is used, with or without the provisions for moving the pole and zero with inductors; but the oven is programmed to dwell at each of a number of fixed temperatures. At each fixed temperature the varactor sweep method is used to search for dips. In this way, very onerous temperature runs in search of dips occurring over extremely small temperature intervals¹³ could be obviated. The C_L in effect interpolates between fixed temperatures.

Acknowledgment

We take this opportunity to thank Mr. T. J. Lukaszek of ECOM for his enthusiastic help; Prof. H. Fukuyo of Tokyo Institute of Technology for responding quickly to our request for reprints and for permission to use Figure 1; Drs. E. Hafner, J. R. Vig, and R. Filler of ECOM for very useful discussions and ideas; Dr. A. Zumsteg of SSIH, Switzerland, for a cogent remark; and Mrs. Anita Earle for her expert typing.

Appendix A

Estimate of Nonlinearities in AT-Cut Resonators

Many of the most important kinds of activity dips arise from nonlinear processes. In order to get a feeling for the magnitudes that might be involved, we evaluate the ratio of the nonlinear to linear terms in the problem of intermodulation solved by Tiersten.^{14,15}

From the nonlinear equation

$$\mathcal{C}_{26}^E u_{1,22} + e_{26} \varphi_{2,22} - \rho \ddot{u}_1 = -\gamma [(u_{1,2})^3]_{,2} \quad (\text{A-1})$$

we use the ratio

$$\mathcal{S} = \left| \frac{-\gamma [(u_{1,2})^3]_{,2}}{\mathcal{C}_{26}^E u_{1,22}} \right| = \frac{3\gamma (u_{1,2})^2}{\mathcal{C}_{26}^E} \quad (\text{A-2})$$

as a measure of the size of the nonlinearity. For the AT-cut,

$$\gamma = 13.7 \times 10^{11} \text{ Pa} \quad (\text{A-3})$$

Making use of the approximate, linear solution for u_1 , evaluated at resonance and at the center of the plate where $u_{1,2}$ is greatest, gives

$$\mathcal{S} = \frac{3\gamma |A_M|^2 \chi^2}{\mathcal{C}_{26}^E} \quad (\text{A-4})$$

where $|A_M| \approx \frac{4}{\pi^2} \cdot \frac{1}{M^2} \cdot \frac{e_{26} V Q_M}{\mathcal{C}_{26}^E} \quad (\text{A-5})$

For AT-cut quartz

$$\mathcal{C}_{26}^E = 29.0 \times 10^9 \text{ Pa}; e_{26} = 0.095 \text{ C/m}^2 \quad (\text{A-6})$$

Taking as an example a 10 MHz fundamental resonator with Q of 500,000 and 1 mV applied voltage gives

$$\mathcal{S} \approx 2.2 \times 10^{-8} \quad (\text{A-7})$$

as the size of the nonlinear term compared to the linear.

Appendix B

Resonator Amplitude with Load Capacitor

In considerations of how far f may realistically be pulled from f_R by additional circuitry, a knowledge of the displacement, stress, and strain fields of the resonator is required. Limiting ourselves to the single shear mode of an unbounded AT-cut plate allows these quantities to be computed simply.

For a plate of thickness $2h$ and applied voltage V we have¹⁶

$$u_1 = \frac{-eh \sin(\frac{\chi_2}{h} X)}{\bar{\mathcal{C}} X \cos X} \cdot a_2 \quad (\text{B-1})$$

$$S_6 = u_{1,2} = \frac{-e \cos(\frac{\chi_2}{h} X)}{\bar{\mathcal{C}} \cos X} \cdot a_2 \quad (\text{B-2})$$

$$T_6 = e \cdot \left\{ 1 - \frac{\cos(\frac{\chi_2}{h} X)}{\cos X} \right\} \cdot a_2 \quad (\text{B-3})$$

$$a_2 = \frac{V/2h}{[1 - k^2 \tan X/X]} \quad (\text{B-4})$$

In the above,

$$e = e_{26}; \bar{\mathcal{C}} = \mathcal{C}_{26}^E + e_{26}^2/\epsilon_{11}^E;$$

$$k^2 = k_{26}^2 = e_{26}^2/(\epsilon_{11}^E \bar{\mathcal{C}}_{26});$$

$$\chi = \frac{\pi}{2} \cdot (f/f_{A0}^{(1)}); f_{A0}^{(1)} = (\bar{\mathcal{C}}/\rho)^{1/2}/4h. \quad (\text{B-5})$$

The displacement u_1 is greatest at $x_2 = \frac{1}{2}h$, while the strain is largest at $x_2 = 0$. In terms of these maximum values we have

$$(u_1/h)_{\text{surface}} = (S_6)_{\text{center}} \cdot \frac{\sin X}{X} \quad (\text{B-6})$$

The resonance frequencies are obtained from the poles of a_2 , that is, from the roots of

$$\tan X = X/k^2 \quad (\text{B-7})$$

About these frequencies $f_{R0}^{(M)}$, the presence of loss serves to limit the amplitude. The amplitude may be found by substituting

$$f_{R0}^{(M)} \cdot (1 + j/2Q_M) \quad (\text{B-8})$$

for $f_{R0}^{(M)}$ in B-1 to B-4.

When C_L is used to pull the frequency to f_L then the effect can be calculated by using

$$k_L^2 = k^2 (1 - \alpha), \quad (\text{B-9})$$

with

$$\alpha = C_0 / (C_0 + C_L), \quad (\text{B-10})$$

in place of k^2 in B-7. This determines f_L . With the known value of f_L inserted into B-1 to B-4, the various amplitudes are determined away from resonance. As long as f_L is not too close to f_R , B-8 need not be used for purposes of determining the effect of a load capacitor.

References*

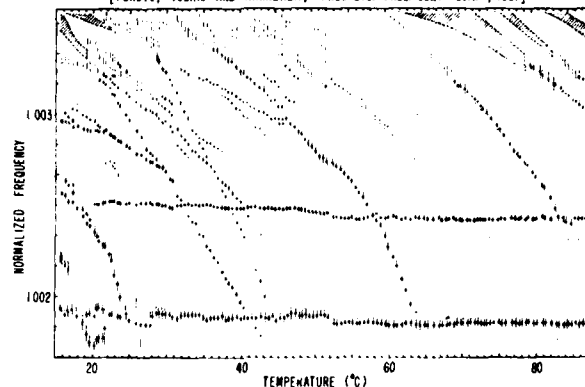
1. E. Hafner; "Some Phenomena in VHF Crystal Units," Proc. 10th AFCS, May 1956, pp. 182-189.
2. Fukuyo, H. Yoshie, and M. Nakazawa; "The Unwanted Responses of Crystal Oscillator Controlled by AT-Cut Plate," Bull. Tokyo Inst. Tech., No. 82, Sept. 1967, pp. 53-64; Proc. 21st AFCS, April 1967, pp. 402-419.
3. C. Franx; "On Activity Dips of AT Crystals at High Levels of Drive," Proc. 21st AFCS, April 1967, pp. 436-454.
4. A. F. B. Wood and A. Seed; "Activity Dips in AT-Cut Crystals," Proc. 21st AFCS, April 1967, pp. 420-435.
5. I. Koga; "Anomalous Vibrations in AT-Cut Plates," Proc. 23rd AFCS, May 1969, pp. 128-131.
6. J. Birch and D. A. Weston; "Frequency/Temperature, Activity/Temperature Anomalies in High Frequency Quartz Crystal Units," Proc. 30th AFCS, June 1976, pp. 32-39.
7. A. Ballato and G. J. Iafrate; "The Angular Dependence of Piezoelectric Plate Frequencies and Their Temperature Coefficients," Proc. 30th AFCS, June 1976, pp. 141-156.
8. A. Ballato; "Doubly Rotated Thickness Mode Plate Vibrators," in "Physical Acoustics: Principles and Methods" (W. P. Mason and R. N. Thurston, eds.), Vol. 13, Chap. 5. Academic Press, New York, 1977.

*AFCS: Annual Frequency Control Symposium, US Army Electronics Command, Fort Monmouth, NJ 07703.

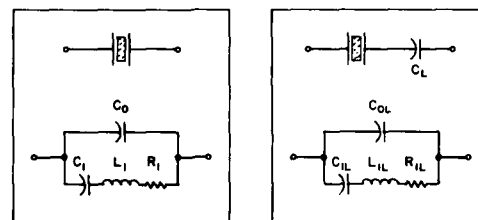
9. E. P. EerNisse; "Calculations on the Stress Compensated (SC-Cut) Quartz Resonator," Proc. 30th AFCS, June 1976, pp. 8-11.
10. J. A. Kusters and J. G. Leach; "Further Experimental Data on Stress and Thermal Gradient Compensated Crystals," Proc. IEEE, Vol. 65, No. 2, Feb. 1977, pp. 282-284.
11. J. A. Kusters; private communication, June 1977.
12. R. D. Mindlin; "Thickness-Shear and Flexural Vibrations of Crystal Plates," J. Appl. Phys., Vol. 22, No. 3, March 1951, pp. 316-323.
13. P. F. Godwin, Jr., and G. L. Snider; "Methods for Production Screening for Anomalous Responses in Quartz Crystals Intended for High Reliability Applications," these Proceedings.
14. H. F. Tiersten; "Analysis of Intermodulation in Rotated Y-Cut Quartz Thickness-Shear Resonators," Proc. 28th AFCS, May 1974, pp. 1-4.
15. H. F. Tiersten; "Analysis of Intermodulation in Thickness-Shear and Trapped Energy Resonators," J. Acoust. Soc. Am., Vol. 57, No. 3, March 1975, pp. 667-681.
16. A. Ballato; "Transmission-Line Analogs for Piezoelectric Layered Structures," Technical Report ECOM-4413, US Army Electronics Command, Fort Monmouth, NJ 07703, May 1976, 252 pp.

MODE CHART OF UNWANTED RESPONSES

[FUKUYO, YOSHIE AND NAKAZAWA, PROC 21st FREQ CONT SYMP, 1967]



1. AT-cut unwanted mode spectrum as function of temperature.



$$f_s = 1/2\pi\sqrt{L_1 C_1}$$

$$r = C_0/C_1$$

$$\alpha = C_0/(C_0 + C_1)$$

$$f_L = f_s \sqrt{1+\alpha/r}$$

$$r_L = r/(1-\alpha)$$

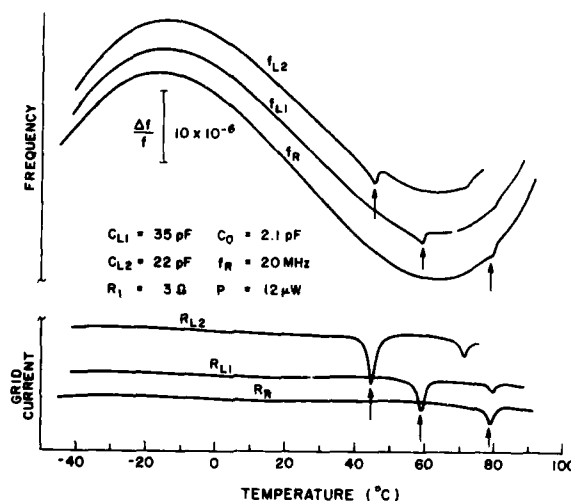
$$C_{0L} = C_0(1-\alpha)$$

$$C_{1L} = C_1(1-\alpha)^2/(1+\alpha/r)$$

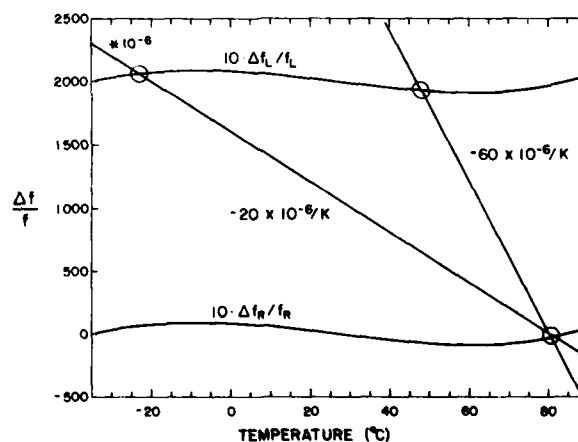
$$R_{1L} = R_1/(1-\alpha)^2$$

$$L_{1L} = L_1/(1-\alpha)^2$$

2. Equivalent circuit parameters of a crystal without and with series load capacitor.

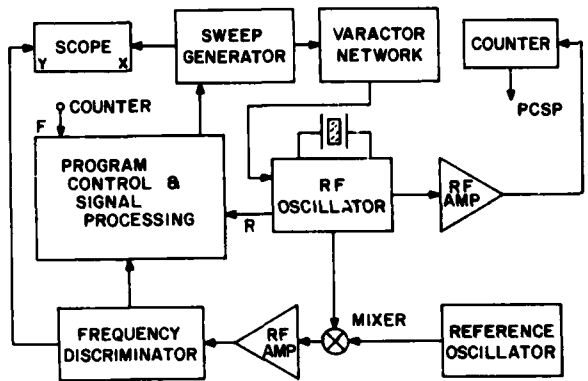


3. Frequency- and activity-temperature plots of a crystal without and with series load capacitors. Locations of activity dips are marked with arrows.

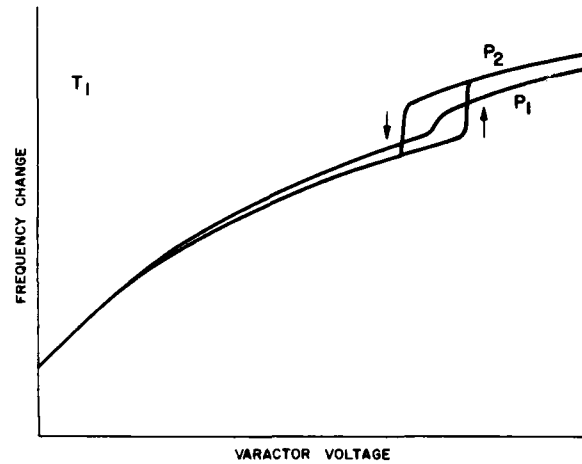


4. Construction for demonstrating change of activity dip temperature with load capacitance.

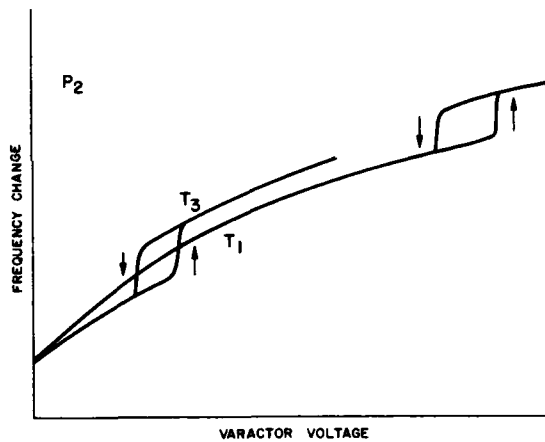
ACTIVITY DIP SCANNER



5. Block diagram of activity dip scanner.

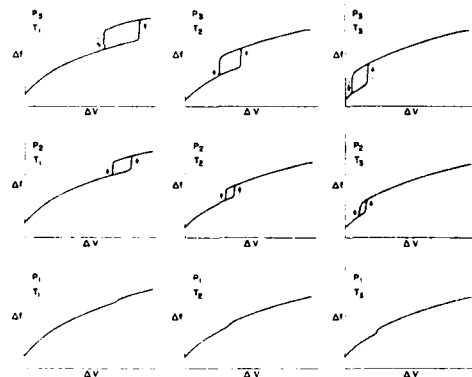


6. Experimental detection of activity dip. Constant temperature; two power levels.

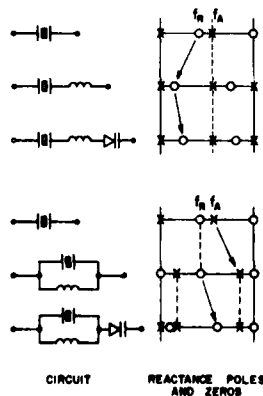


7. Experimental detection of activity dip. Constant power level; two temperatures.

FREQUENCY CHANGE VERSUS VARACTOR VOLTAGE AS FUNCTION OF TEMPERATURE AND POWER LEVEL



8. Montage of experimental plots of frequency change versus varactor voltage. Various power level and temperature combinations.



9. Alteration of crystal pole-zero locations by addition of circuit elements.

MEASUREMENT OF THE CHARACTERISTIC

FREQUENCY OF AN AT-CUT PLATE

John H. Sherman, Jr.
GENERAL ELECTRIC COMPANY
Lynchburg, Virginia 24502

Abstract

The measurement method of Atanasoff and Hart is considered in terms of the familiar Sykes-McSkimin and Butterworth-Van Dyke models of the crystal resonator. After consistency is established, the criterion is extracted for a measurement of resonant frequency in an air gap to yield a value equal to the characteristic frequency. The capacitance ratio of the unelectroded blank is computed and air gap design formulas are written with numerical values. The relation between the measurements on the overtones is discussed with illustration of advantages sometimes found in overtone measurement. Laboratory and factory measurement objectives are compared. A practical test fixture is described and illustrated.

Key Words: Measurement, frequency, characteristic frequency, blank, monolithic crystal filters, MCF, coupling, trapped energy.

Introduction

A successful strategy for the design of monolithic crystal filters consists of the following, only slightly simplified steps:

- 1) Based on the required filter performance, select an appropriate normalized filter prototype.
- 2) Based upon the terminating impedance, center frequency and bandwidth required, denormalize the filter. Determine the element inductance and the inter-element coupling coefficients.
- 3) From the element inductance and center frequency, determine the electrode area. Design an element electrode.
- 4) Using the electrode dimensions chosen in 3 and the coupling coefficients found in 2, plot for each coefficient curves of plateback vs separation along X and Z.
- 5) Choose the axis along which the array will be coupled and the single value of plateback which provides the most reasonable set of element separations along the chosen axis.
- 6) Add the chosen value of plateback to the filter center frequency to obtain the characteristic frequency of the wafer.
- 7) Test the overall array for unwanted responses in the design [1].

To implement the design, dimensions are established for the periphery of the wafer and fixtures made. Then blanks are prepared at a number of frequencies in the neighborhood of the computed characteristic frequency so sample filters can be made, in order to determine the actual measured frequency which must be attained to realize the filter objective.

The empirical determination of blank frequency has to be made because the characteristic frequency of the blank is generally not observed in a blank frequency measurement. What is usually observed is the frequency of a crystal controlled oscillator in which the unelectroded crystal wafer is positioned in an electrode structure of distressingly incon-

stant configuration in an oscillator of similarly undisciplined characteristics. The practical measurement problems become so evident that extreme measures are often taken to beef up the electrode structure and define the oscillator circuitry. Even so, repetitive measurement of a given blank regularly results in a large spread of measurements apparently dependent upon such variables as which side of the wafer is up.

This inaccessibility of the characteristic frequency for measurement is more basic than simply a difficult problem in fixturing for measurement. It is shown below that the best model we use for the resonant frequency of a finite wafer makes explicit that the frequency we expect to observe is not the characteristic frequency itself, but something quite other. Only under unique circumstances is the characteristic frequency observable.

This has been a stumbling block for some time. The characteristic frequency enters the design of "clean spectrum" or single-mode energy trapped resonators in exactly the same way as it enters monolithic crystal filter design. A major factor in our inability to reach a consensus on the values of the "magic numbers" in the plateback formulas [1], [2], [3] used in practical resonator design is simply that the characteristic frequency of the wafers fabricated is not known. This has made choice between the various proposals of theoretical maximum values of plateback not subject to test. Since theory leads to a maximum only regardless of the details, the industry has been able to exploit the theory by working to practical limits treated as less than the absolute maximum, in which the measurement uncertainty is absorbed into an empirically reduced numerical coefficient.

A strategy of measurement which will yield a set of readings of small spread is necessary. One which will yield the characteristic frequency itself will be of signal additional benefit.

This paper treats the search for such a strategy.

The Three-Dimensional Resonant Piezoid

Atanasoff and Hart [4] in their 1941 paper observe that "... no easy exact theory of the vibrations of a finite plate is possible ...". Their measurement strategy was to use very high overtones. The argument from physical intuition which led to this strategy was that the more nodal surfaces were stacked between the major surfaces, the more closely those nodal surfaces would have to lie in planes and the more nearly they would have to be equally spaced, parallel to each other and parallel to the major surfaces. Koga had already showed that when the phase surfaces of the acoustic standing wave are parallel to the surfaces of the plate the overtones are integrally related. By measuring on very high overtones (up to the 87th in thickness and the 263rd laterally) they were able to demonstrate that the

ratio of the frequency to the overtone order became a constant. They took this constant to be the characteristic frequency.

At about the same time Sykes [5] and McSkimin [6] were developing their powerful expression for the frequencies of a rectangular prism of quartz in body shear. This formula, published just two years after the Atanasoff and Hart paper, has given results which are found quite accurate. While the formula is not exact, it is remarkably accurate and quite easy. It is readily shown to be consistent with the Atanasoff and Hart method and result at high overtones, while its principal and well-tested application has been at low orders of overtone and enharmonic index.

The Sykes-McSkimin equation for the antiresonant frequencies of an At-cut crystal plate is:

$$f_{mnp} = \frac{1}{2\sqrt{\rho}} \sqrt{\frac{m^2 \bar{C}_{66}'}{t^2} + \frac{n^2 \bar{C}_{11}'}{X^2} + \frac{p^2 \bar{C}_{55}'}{Z^2}} \quad (1)$$

where X and Z are the edge dimensions of the plate along the x and z' axes respectively and t is the thickness of y' dimension. Simple algebraic manipulation yields:

$$f_{mnp} = \frac{m}{2t} \sqrt{\frac{\bar{C}_{66}'}{\rho}} \sqrt{1 + \frac{t^2}{m^2 \bar{C}_{66}'} \left(\frac{n^2 \bar{C}_{11}'}{X^2} + \frac{p^2 \bar{C}_{55}'}{Z^2} \right)} \quad (2)$$

This is identified to be

$$f_{mnp} = mf_{100} \sqrt{1 + \frac{t^2}{m^2 \bar{C}_{66}'} \left(\frac{n^2 \bar{C}_{11}'}{X^2} + \frac{p^2 \bar{C}_{55}'}{Z^2} \right)} \quad (3)$$

which is confirmed by substituting 1, 0, 0 for m, n, p respectively in (1).

The frequency f_{100} is the characteristic frequency sought. The measured frequency is the frequency resulting from the interaction of the wafer of antiresonant frequency f_{mnp} and the measuring circuit. It is a simplification even to look on f_{mnp} as accurately measurable. Nevertheless, in order to do things in a reasonable order and keep ideas simple, it is useful to treat f_{mnp} as accessible. In any case:

$$f_{100} = f_{mnp} / m \sqrt{1 + \frac{t^2}{m^2 \bar{C}_{66}'} \left(\frac{n^2 \bar{C}_{11}'}{X^2} + \frac{p^2 \bar{C}_{55}'}{Z^2} \right)} \quad (4)$$

and we can approximate f_{100} by f_{mnp}/m .

The fractional error in this approximation is itself approximated by

$$\delta_1 = \frac{-t^2}{2m^2 \bar{C}_{66}'} \left(\frac{n^2 \bar{C}_{11}'}{X^2} + \frac{p^2 \bar{C}_{55}'}{Z^2} \right) \quad (5)$$

which must be minimized. Stated simply, numerators are to be made small and denominators are to be made large in order to minimize the error. The condition $n = p = 0$ is the Koga condition of planar phase surfaces parallel to the major surface of the wafer. This cannot occur in a finite wafer; the minimum value of n is unity in a practical wafer. Recast for emphasis, (5) becomes

$$\delta_1 = \frac{1}{2} \left(\frac{n^2 t^2 \bar{C}_{11}'}{m^2 X^2 \bar{C}_{66}'} + \frac{p^2 t^2 \bar{C}_{55}'}{m^2 Z^2 \bar{C}_{66}'} \right) \quad (5a)$$

and the justification of the measurement at high overtones is displayed, extracted from the Sykes-McSkimin equation.

To make f_{mnp}/m be as accurate an estimate as possible of f_{100} involves three independent elements:

- 1) Make m as large as possible by measuring on the highest practicable harmonic mode.
- 2) Make n and p as small as possible. This means to measure the lowest frequency member of the enharmonic sequence associated with the overtone. This is usually what is measured in a naive setup, but pitfalls exist. The resonator should be swept and the lowest response of the sequence identified, even if it has only a small response.
- 3) Make the X and Z dimensions of the blank be as large as possible. In principle this means to perform final preparation of the major faces of the wafer prior to final dimensioning.

Measurement of an Unelectroded Wafer

The accepted equivalent circuit for the sequence of "main" responses of a crystal is given:

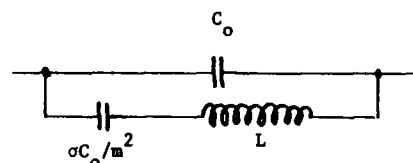


Fig. 1

where σ is written for the reciprocal of the conventional ratio of the capacitances, r . This circuit correctly displays that the capacitance ratio varies as the square of the overtone order. It does not, however, display correctly that the harmonic sequence related by integer ratios is of the antiresonant frequencies. This circuit model implies incorrectly that the series resonant frequencies are harmonically related. The model does, account correctly for the spacing of the series and antiresonant frequencies. It turns out that this is exactly what we require of the model.

When a crystal plate is operated in an air gap holder, the air gap constitutes a parallel plate condenser in series with the crystal, leading to this equivalent circuit for a blank in a test fixture:

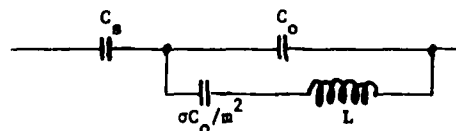


Fig. 2

The antiresonant frequency of the wafer is still the focus of attention, but is inaccessible for measurement with practical equipment. At the blank's antiresonant frequency the impedance of the circuit is high. Crystal impedance meters and pi section transmission sets look at the low impedance condition of the equivalent circuit. Other oscillator types look at some intermediate impedance condition and at a frequency between these two frequencies.

The equivalent circuit of Figure 1 has resonant frequencies:

$$f_{sm} = \frac{m}{2\pi} \sqrt{\frac{1}{oLC_o}} \quad (6s)$$

$$f_{am} = f_{sm} \sqrt{1 + \frac{\sigma}{m^2}} \quad (6a)$$

The equivalent circuit of Figure 2 has resonant frequencies:

$$f_{rm} = f_{sm} \sqrt{1 + \frac{\sigma C_o}{m^2 (C_o + C_s)}} \quad (7s)$$

$$f_{am} = f_{sm} \sqrt{1 + \frac{\sigma}{m^2}} \quad (7a)$$

The frequency f_{am} is seen to be the antiresonant frequency of the blank both in and out of the gap. We wish to consider f_{am} while f_{rm} is available for ready measurement. The ratio of these is

$$\frac{f_{am}}{f_{rm}} = \sqrt{\frac{1 + \frac{\sigma}{m^2}}{1 + \frac{\sigma}{m^2 (1+\gamma)}}}, \quad \gamma = \frac{C_s}{C_o} \quad (8)$$

which with manipulation reduces to

$$\frac{f_{am}}{f_{rm}} = \sqrt{1 + \frac{\gamma}{1+m^2(1+\gamma)/\sigma}} \quad (9)$$

Using the measured frequency f_{rm} as an approximation of the wafer's antiresonant frequency has a fractional error of approximately

$$\delta_2 = \frac{\gamma}{2[1+m^2(\gamma+1)/\sigma]} \quad (10)$$

For small γ and large m this approaches

$$\delta_2 \rightarrow \frac{\sigma \gamma}{2m^2}$$

The error in determination of f_{am} incurred by measuring f_{rm} is minimized by measuring on the highest practicable overtone. Since the error is nearly proportional to γ , minimizing γ minimizes the error.

C_s must be made small. In terms of structure of the test fixture, the air gap must be the maximum practicable length.

Naively one might be tempted to try a small electrode in order to decrease γ . While it is true that this would decrease C_s , it would not decrease γ as the value of C_o would also be decreased. C_o is the capacitance of that portion of the blank excited by the electric field of the electrodes. The air gap electrodes should remain sensibly the size of the wafer. Bridge circuitry can be devised to use electrodes much larger than the blank without loss of accuracy or sensitivity. Perhaps a theory of measurement of equal power can be developed for an electrode of small (vanishing) area. Such is not attempted in this study.

So far this study has looked at the procedure of Atanasoff and Hart in terms of models not used by them and verified that these models afford their method the validity they postulated and demonstrated for it. In the process an algebraic model has been created and in that algebraic model is buried a happy surprise.

Direct Measurement of the Characteristic Frequency

The frequency of a plate is always higher than the characteristic frequency. The relationship is derived from (1)

$$\frac{f_{mnp}}{f_{moo}} = \sqrt{1 + \frac{t^2}{m^2 \bar{C}_{66}'} \left(\frac{n^2 \bar{C}_{11}'}{x^2} + \frac{p^2 \bar{C}_{55}'}{z^2} \right)} \quad (11)$$

The principal resonance of a blank is identified by Sykes [5] to be f_{m10} . Van Dyke [7] illustrates the 110 mode of a rectangular plate in a 3-dimensional map of measurements made on an actual plate. Thus the principal mode is given by:

$$\frac{f_{m10}}{f_{m00}} = \sqrt{1 + \frac{t^2 \bar{C}_{11}'}{m^2 x^2 \bar{C}_{55}'}} \quad (12)$$

The resonant frequency of the combination of blank and air gap test fixture is always lower in frequency than the frequency of blank (anti) resonance. The relationship is given by:

$$\frac{f_{am}}{f_{rm}} = \sqrt{1 + \frac{\gamma}{[1+m^2(1+\gamma)/\sigma]}}, \quad \gamma = \frac{C_s}{C_o} \quad (9)$$

Remembering that f_{am} and f_{m10} are both symbols in their own models for the antiresonant frequency of the finite plate on the lowest member of the sequence of enharmonics associated with the m -th thickness mode, it is recognized that f_{am} and f_{m10} are identically equal. It is furthermore significant that the characteristic mode frequency $f_{moo} = mf_{100}$ is lower in frequency than f_{m10} and that the measured frequency f_{rm} is lower in frequency than f_{am} .

If we divide (12) by (9) there results:

$$\frac{f_{m10}}{f_{moo}} = \frac{f_{rm}}{f_{moo}} = \frac{1 + \frac{t^2 \bar{C}'_{11}}{m^2 X^2 \bar{C}'_{66}}}{1 + \frac{\gamma}{1+m^2(1+\gamma)/\sigma}} \quad (13)$$

and if it is possible to create in a practical measurement situation conditions to make the radical evaluate to unity, then the measured frequency f_{rm} will equal f_{moo} , the desired mode characteristic frequency.

The radical becomes unity if the increments of the numerator and denominator are equal, i.e., if:

$$\frac{\gamma}{1+m^2(1+\gamma)/\sigma} = \frac{t^2 \bar{C}'_{11}}{m^2 X^2 \bar{C}'_{66}} \quad (14)$$

Solved explicitly for γ , (14) yields:

$$\gamma = \frac{1 + \frac{\sigma}{m^2}}{\frac{\sigma X^2 \bar{C}'_{66}}{t^2 \bar{C}'_{11}} - 1} \quad (15)$$

Defining:

$$\gamma_o = \left(\frac{\sigma X^2 \bar{C}'_{66}}{t^2 \bar{C}'_{11}} - 1 \right)^{-1} \quad (16)$$

(15) may be rewritten:

$$\gamma = (1 + \sigma/m^2) \gamma_o \quad (17)$$

Thus, to the accuracy of our model, when this value of γ is achieved in a practical test configuration the characteristic frequency can be measured directly.

Application to the Design of Air Gap Fixtures

An air gap fixture is made with a fixed gap g which in operation contains a crystal of thickness t with a residual air gap length ℓ , so that

$$g = \ell + t$$

The capacitance of the blank is given by the formula

$$C_o = \epsilon_o^S A/t$$

while that of the gap is

$$C_g = \epsilon_o A/\ell$$

Since the areas of the electrodes and the blank are equal,

$$\gamma = C_g/C_o = \epsilon_o t/\epsilon_{22}^S \ell$$

The ratio $\epsilon_{22}^S/\epsilon_o$ is known as the "specific inductive capacity" of relative dielectric constant of quartz, ϵ . Hence

$$\gamma = t/\epsilon \ell \quad (18)$$

and the gap becomes

$$g = \ell + t = t(1 + \epsilon \gamma) \quad (19)$$

In terms of the blank dimensions

$$g = t \left(1 + \frac{\frac{\sigma X^2 \bar{C}'_{66}}{t^2 \bar{C}'_{11}} - 1}{\epsilon(1 + \sigma/m^2)} \right) \quad (20)$$

Numerical Values

Only a few constants are required to convert these relationships to design equations.

The elastic constants are the appropriately rotated and stiffened constants. The values for AT-cut quartz are:

$$\bar{C}'_{11} = 110.602(10)^9$$

$$\bar{C}'_{66} = 29.350(10)^9$$

The relative dielectric constant is not truly a constant, as quartz has different dielectric constants along and perpendicular to Z. The entire range is only some 2%. The entire family of body shear crystals with cubic temperature-frequency characteristics is found in a band of θ between 33.25° and 35.5° and the range of dielectric constant for this entire class of cuts is less than 0.075% wide, which compares closely with the accuracy of the best value of any constant known for AT-cut quartz. The entire range of ϵ is calculated to lie between 4.497 and 4.500. The value 4.500 will be used for ϵ .

The ratio of the capacitances is determined by a method due to Bechmann as exploited by Sauerbrey [8]. Starting with a known or assumed amplitude distribution across the entire resonator one performs certain integrations and computes a motional capacitance. The dependence motional capacitance on enharmonic order is thus displayed, and of the capacitance ratio. The mode of resonance has indices 1, 1, 0 which is shorthand for a distribution which varies as a half cycle of a cosine function along X and is constant along Z.

Performing the integrations one extracts a ratio

$$r_{c110}/r_{c100} = \frac{\pi^2}{8}$$

which combines with a best value for r_{c100} of 159.4 to produce a value for r_{c110} of 196.7. The reciprocal

of this value is what we have been calling σ

$$\sigma = 5.084(10)^{-3} \quad (21)$$

and

$$\frac{\bar{C}'_{11}}{\sigma \bar{C}'_{66}} = 741.2 \quad (22)$$

Since the frequency constant for the characteristic frequency is $1.660(10)^6$ Hz mm the X/t ratio is calculable by the formula

$$\frac{X}{t} = \frac{fX(10)^{-6}}{1.660} \quad (23)$$

where f is the frequency in Hz and X is the x dimension in millimeters. If X is measured in inches then the frequency constant is $65.35(10)^3$. By this relation the X/t ratio is obtainable to an accuracy of four digits or to the accuracy of measurement of X (whichever is less) even though direct measurement of t to this kind of accuracy is impossible.

Using these numerical values the formula for γ_0 becomes:

$$\gamma_0 = \frac{1}{\frac{X^2}{741.2t^2} - 1} \quad (24)$$

and

$$\gamma = 1 + \frac{5.084(10)^{-3}}{m^2} \gamma_0 \quad (25)$$

A plot of (24) is provided as part of Figure 3. Solutions exist for values of X/t greater than $\sqrt{741.2}$ so for blanks of reasonably conventional proportions it is quite possible to construct an air gap fixture in which the frequency measured is the characteristic frequency.

Also included in Figure 3 are plots of

$$l/t - 1/\epsilon\gamma = 1/4.5\gamma \quad (26)$$

and

$$g/t = (l+t)/t = 1 + 1/4.5\gamma \quad (27)$$

as well as the reciprocal of (27) called f'/f:

$$f'/f = f.5\gamma/(1+4.5\gamma). \quad (28)$$

The significance of the ratio f'/f is that in gauging an air gap it is extremely practical to use a quartz blank. The ratio f'/f is just the ratio of the characteristic frequency of the gauge blank to that of the blank to be measured. The gauge blank has a lower frequency than the work, hence a greater thickness, a lesser X/t ratio and a smaller air gap for its measurement.

Using an uncalibrated air gap holder to measure a crystal blank which will be used in turn to gauge a critically adjusted air gap sounds risky. It is a procedure which has to be defended by its necessity. In almost no other conceivable way can we generate the number of precise gauge blocks we require to have the flexibility to respond to commercial orders. We

have to be able to make the system work. Furthermore, when working in actual numbers and real dimensions in millimeters or inches we only know numbers to four digits. On this basis we cannot do better than several hundred parts per million in any number, including f'. While it is not a foregone conclusion that a measurement of a blank of f' will be within 500 ppm of its true characteristic frequency, we are discussing how to try to make it so.

To clarify the matter further, Figure 4 displays the way the characteristic frequency falls between f_{g1} and f_{a1} while Figure 6 (which can be obtained from Figures 3 and 4) displays the way the measured frequency f_1 falls between the same limits. These limits for the blank are separated by 2540 ppm. From these curves guidance can be found properly to proportion the gap to measure the gauge blank to the requisite accuracy.

"Exact" and Approximate Forms

Because these various constants are known for quartz only to four digits accuracy, no direct numerical value can be calculated to better than four digits. Careful examination of the limitation of significant digits shows, for example, that the calculated values of γ greater than 1 can have only three significant digits. The entire plot of Figure 3 is in no more than three digits of detail and the multiplier $(1+\sigma/m^2)$ of (17) differs from unity only in the fourth digit at most. Thus to the accuracy of the plot γ and γ_0 are indistinguishable and Figure 3 can contain no information concerning the overtone. For this reason the various curves of the figure are not subscripted with a naught though the algebraic expression of (16) must be.

If in (14) attention is directed to the denominator of the left term:

$$\frac{\gamma}{1 + m^2(1+\gamma)/\sigma} = \frac{t^2 \bar{C}'_{11}}{m^2 X^2 \bar{C}'_{66}} \quad (14)$$

it is seen that on the fundamental the denominator will be roundly 200 or more, on the third overtone some 2000, and on the fifth overtone some 5000. Going up at the square of the overtone, the seventh would bring it to 10000 and the ninth over 16000. The poor lonely additive 1 in that denominator is on the fundamental only 1/2% and rapidly shrinks as the overtone order increases. If it is therefore disregarded (14) can be approximated:

$$\frac{\gamma\gamma}{m^2(1+\gamma)} = \frac{t^2 \bar{C}'_{11}}{m^2 X^2 \bar{C}'_{66}} \quad (29)$$

When this is solved for γ there results:

$$\gamma = \frac{t^2 \bar{C}'_{11}}{X^2 \bar{C}'_{66} - \frac{t^2 \bar{C}'_{11}}{X^2 \bar{C}'_{66}}} \quad (30)$$

which reduces with a little more manipulation to (16). Thus from (29) we conclude that $\gamma = \gamma_0$ and clarify

somewhat the significance of the multiplier in (17). To at least a first approximation the relation is independent of overtone order and the measured frequencies on the overtones should be very nearly related as the overtone orders.

If the approximate value γ_0 is substituted for γ in (13) the sensitivity of (13) to variation of γ can be estimated without differentiation. (Remember if the complete value of γ is inserted, (13) reduces identically to unity.) This substitution is the model of measuring a blank on a low overtone in an air gap critically set for high overtones. When the substitution is made there results:

$$\frac{f_{rmo}}{f_{mo}} = \sqrt{\left(1 + \frac{t^2 \bar{C}_{11}'}{m^2 X^2 \bar{C}_{66}'}\right) \left[1 - \left(\frac{1}{m^2 + \sigma}\right) \left(\frac{t^2 \bar{C}_{11}'}{X^2 \bar{C}_{66}'}\right)\right]} \quad (31)$$

or, numerically

$$\frac{f_{rmo}}{f_{mo}} = \sqrt{\left(1 + \frac{3.768}{m^2} \frac{t^2}{X^2}\right) \left[1 - \frac{3.768}{(m^2 + 0.005)} \frac{t^2}{X^2}\right]} \quad (32)$$

The maximum difference between γ and γ_0 occurs on the fundamental where the difference is just about 1/2 percent and the maximum difference between f_{rmo} and f_{mo} occurs likewise on the fundamental. Additionally it occurs at the smallest value of X/t which was shown earlier to evaluate to $\sqrt{741.2}$. When (32) is evaluated on the fundamental and at $X/t = 27.22$ the result is obtained:

$$\begin{aligned} \frac{f_{r10}}{f_{100}} &= \sqrt{\left(1 + \frac{3.768}{741.2}\right) \left[1 - \frac{3.768}{1.005(741.2)}\right]} \\ &= 1 - 2.116(10)^{-7} \\ &= .9999997884 \end{aligned}$$

If the air gap is set for γ_0 instead of γ , the error in reading f_{100} is only about -0.2 ppm. The error in f_{300} is two orders of magnitude less. Furthermore, the ratio of the frequency measured on the third to that measured on the fundamental will differ from 3 by -0.2 ppm. And this is the worst case. For all greater values of X/t the difference is even less. Now the final statement of the third paragraph above can be made much stronger. In any case of practical measurement in a properly dimensioned air gap the overtones will be, to the accuracy attainable, related as the overtone orders.

However accurate or inaccurate the gauge blank is can be evaluated by the ratio of the measured frequencies on the fundamental and third overtone. With excellent blanks in a proper gap the ratio should be 3. With this criterion at hand, gauge blanks can be adjusted with the necessary sensitivity, repetitively setting the gap with the gauge, reading the crystals on the fundamental and the third, then etching the gauge.

Experimental Investigation

Confirmation of this development into hardware has no practical problem of overwhelming substance, but there is a problem of intellectual import. Ideally to verify the accuracy of an instrument one would measure a known sample, a "standard", and thus

calibrate or verify the instrument. The only access other than this to the characteristic frequency known to me is through the laborious method of Atanasoff and Hart, a process much more difficult to carry out on a plate only about 0.15 mm thick than on their blocks some 50 times thicker. In the absence of a blank of known characteristic frequency to measure, confirmation was sought in other peculiar phenomena predicted in the derived relationships. Blanks are available at 9.5 and 11.3 MHz with the same major dimensions, and air gap fixtures have been made to accommodate them. Measurements on these blanks using adaptations of those fixtures yielded results tending to confirm the theory.

A half-lattice was constructed in which the coil was a tunable autotransformer with an air gap in one arm and a balancing capacitor in the other (Figure 5). This device was awkward in the extreme to adjust, but it was possible to balance at 9.5 and at 34 MHz, though not simultaneously. Fifth and seventh overtone response are to be seen in blanks but so far lattices well enough behaved to give strong responses at these overtones have not been realized.

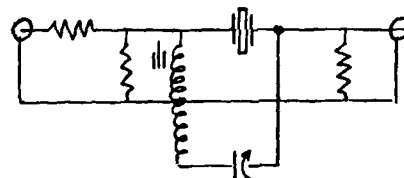


FIGURE 5

Tunable Half-Lattice Fixture

Measurements were made in the HP 8553B spectrum analyzer. The entire observable spectrum of the enharmonic family of interest is displayed in the oscilloscope presentation. In this way any anomalies are displayed and the quality of the measurement evaluated. The marker built into the generator portion allows any particular characteristic of the trace to be isolated and measured with indicated 10 Hz resolution.

Blanks were obtained at 9.500 MHz and 11.300 MHz having an x -dimension of $0.449'' \pm 0.001''$.

When this scanty amount of information is combined with (23), (24), (26) and (27) we obtain:

f	X/t	t	γ_0
9.500	65.42	$6.879(10)^{-3}$	0.2095
11.300	77.81	$5.783(10)^{-3}$	0.1395
		l	f'
		$7.297(10)^{-3}$	4.610
		$9.213(10)^{-3}$	4.358

That the gap is greater for the higher frequency than for the lower may come as a surprise. (It did to me.) Certainly l must increase with frequency since l becomes zero at $X/t = \sqrt{741.2}$. At greater frequencies the X/t ratio is greater than the critical value and l is finite and monotonic. The total gap, the sum of l and t also increases.

The gap was set with a crystal naively measured at 4.515 MHz. This made the gap be long for the 9.5 MHz blanks, but short for the 11.3 MHz. In consequence the load capacitance of the 9.4 MHz blanks, i.e., the gap capacitance C_g , is too low and the fundamental frequency will be measured high. The third overtone frequency will be measured high also but by a lesser amount so that the ratio of the third frequency to the fundamental will be less than 3. Sixteen of the 9.5 MHz crystals measured yielded an average ratio of 2.998924 (σ of 0.000653). Conversely the gap is short for 11.3 MHz. C_g will be large and the fundamental frequency will be measured low. The ratio of the frequency on the third to the frequency on the fundamental should exceed 3. Fourteen of the 11.3 MHz crystals measured yielded an average ratio of 3.001473 (σ of 0.004296).

These measurements were made with the fixture carefully balanced at each measurement. The circuit as realized could be optimized at each frequency but across the range only a compromise could be effected. When the ostensible best compromise was effected the ratios measured in the fixture were all below 3. This continues to be the case in all of the more convenient embodiments of the bridge which have been devised.

The results of this careful set of measurements are taken as constituting a substantial confirmation of the validity of the design equations for the air gap.

The original intent was to etch the 9.5 MHz blanks and measure the ratio as the frequency of the blanks approached that for which the gauge blank was correct, where the ratio would be exactly 3. This became a frustrated hope as the relative activity of interfering modes increased with the increased smoothness of the surface with heavy etching. Figures 6 through 11 illustrate spectra on the fundamental and third overtones showing the drastic effect of interfering modes on the fundamental spectrum as compared with the overtone. In the etching the principal fundamental frequency passed ten even order interfering flexure overtones from 80 through 98 [5]. Many other interfering modes, not identified, are apparently present as well. Too many measurements made at the fundamental were contaminated by interfering modes for the ratio to be observably moving monotonically toward 3.

This blank dimension, originally chosen for 10.7 MHz, encounters an interfering flexure close by the final etch frequency at both of these used frequencies, causing persistent in-process measurement problems. The spectra displayed are of a crystal ready for electroding. Although there is a real problem in deciding where the fundamental resonance occurs, there is no problem at all about the third; the first response is clearly displayed. This is no doubt due to the extremely high overtone order of the interfering modes (now nearly the 300th) where the Q has fallen so low as to have no detectable effect. The resonance peak is only a third as wide as the fundamental, making the determination of its maximum absolutely 3 times and relatively 9 times as precise. As a consequence these blanks, so hard to read on the fundamental, are readily read on the third.

After a number of attempts a final most practical circuit configuration was adopted as shown in figures 12 and 13. The resistor pads terminating the bridge allow the bridge to be balanced at 35 MHz, which is adequate for these measurements. The grounded

gap electrode is the bottom one, allowing the fixture to be used on a metal bench top or on wood without effect on the balance. The circuit was wired on a board laid out for universal application with DIP integrated circuits. The excess conductors introduce stray capacitance and narrow the band over which the bridge is balanced. A special board would permit the same quality of balance to even higher frequency.

Some Final Practicalities

To this point the entire discussion has been directed to making the characteristic frequency observable. This is an undertaking of more significance in the development laboratory than in a strictly production environment. The production measurement must be readily made and reproducible, but does not have to be of the characteristic frequency itself. An offset once determined can relate the readily made reading to the designed frequency.

Thus it is possible with a single adjustment of the bridge to measure the 9.5 and 11.3 MHz blanks on their third overtone with discrepancies of only about 5 ppm at both frequencies referred to the reading with the fixture setup most perfectly.

Even more simplification is possible. Returning to the original ideas at the beginning of the paper, the tuning curve with air gap length can be sketched as shown in Figure 14. The three identified frequencies f_c , f_a and f_s are the characteristic frequency, the Sykes-McSkimin or antiresonant frequency and the series resonant frequency, respectively, of a blank. The curve displays the series resonant frequency of the crystal and gap as the gap grows without limit. The theoretical development of this paper dealt with the identification of the condition where this frequency equals the characteristic frequency.

If the criterion can be relaxed to demand only that the reading be highly reproducible, the figure illustrates that the sensitivity of the reading to errors in the setting of the gap decreases with increasing gap length. The reproducibility of measurement of a frequency above the characteristic frequency is superior to the characteristic frequency itself. Of course the coupling is reduced by increasing the gap and at some gap length becomes too small. The practical consideration is to use the longest gap circumstances will allow. This, combined with the 9 times improvement in using the third overtone, allows very reproducible readings to be made quite readily.

With the spectrum analyzer the gap may be 50 or more times the crystal thickness and still the crystal be read with ease. Additionally the spectrum analyzer is at least 1 1/2 times as fast as the crystal impedance meter. This equipment is not commonly found in crystal factories, however, and it is not the intent to give the impression that it is indispensable. In most instances third overtone readings made in the crystal impedance meter with the longest gap which will allow oscillation will be as reproducible as the production shop will find needed. The bridge network, excited like any other transmission test set, allows the direct calibration of the offset of the oscillating measurements.

Conclusion

The characteristic frequency of a rectangular AT-cut blank has been made accessible to measurement. Simultaneously the practical blank measurement problem

has been illuminated leading to proposals of method which can simplify or improve the reproducibility of production measurements.

References

- [1] Sherman, J. H., "Dimensioning Rectangular Electrodes and Arrays of Electrodes on AT-Cut Quartz Bodies," Proc. 30th Annual Frequency Control Symposium, 1976, pp 54-64.
- [2] Curran, D. R., and Koneval, D. J., "Factors in the Design of VHF Filter Crystals," Proc. 19th AFCS, 1965, pp 213-268.
- [3] Tiersten, H. F., "Analysis of Trapped Energy Resonators Operating in Overtones of Coupled Thickness-Shear and Thickness-Twist," Proc. 29th AFCS, 1975, pp 71-75.
- [4] Atanasoff, J. V. and Hart, P. J., "Dynamical Determination of the Elastic Constants and Their Temperature Coefficients for Quartz," Physical Review, Vol. 59, Jan 1941, pp 85-96.
- [5] Sykes, R. A., "Modes of Motion in Quartz Crystals, the Effects of Coupling and Methods of Design," BSTJ Vol. XXII, No. 1, reprinted as Heising, "Quartz Crystals for Electrical Circuits," Van Nostrand, Chapter 6. (Out of print.)
- [6] McSkimin, H. J., "Theoretical Analysis of Modes of Vibration for Isotropic Rectangular Plates Having All Surfaces Free," BSTJ Vol. XXIII, No. 2, reprinted as Heising Chapter 7 (See [5]).
- [7] Van Dyke, K. S., "The Piezoelectric Survey of Strain Patterns in Thickness Shear Quartz Resonators," Proc. 10th AFCS, pp 1-9. Also "Strain Patterns in Thickness-Shear Resonators," Proc. 11th AFCS, pp 41-61.
- [8] Sauerbrey, G., "Measurement of Amplitude Distributions of Vibrating AT-Cut Crystals by Means of Optical Observations," Proc. 17th AFCS 1963, pp 28-50.

Illustrations

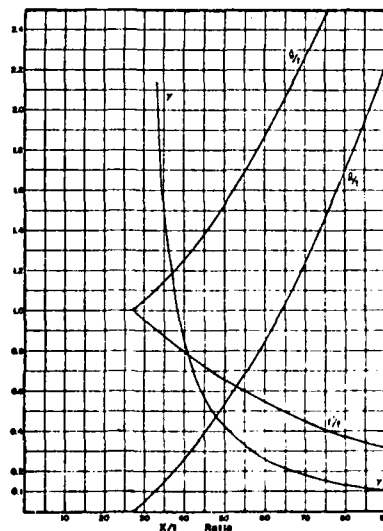


Figure 3: Air Gap Design Parameters.

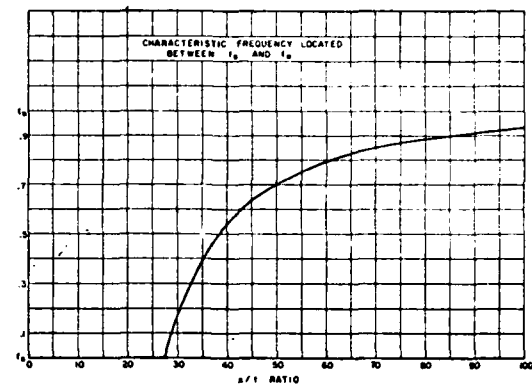


Figure 4: Characteristic Frequency Interpolated Between f_{s1} and f_{a1} .

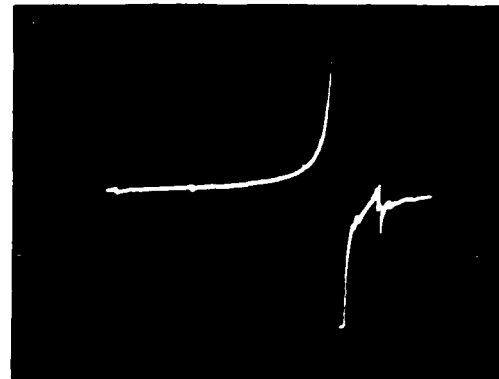


Figure 6: Spectrum illustrating interfering flexure modes. No. 84 interferes with the fundamental main response while 82 is visible to the left. The other mode shown is unidnetified.

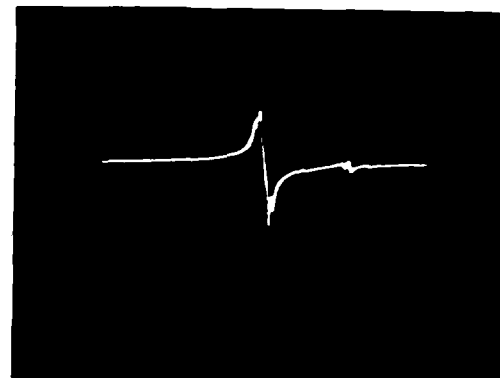


Figure 7: Expanded spectrum of the main response of the blank in Figure 6.

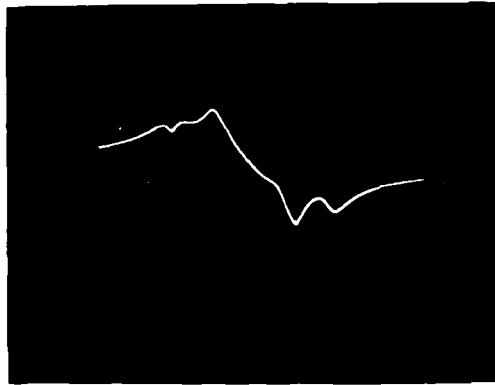


Figure 8: Further expansion showing complexity where the main response peak belongs.

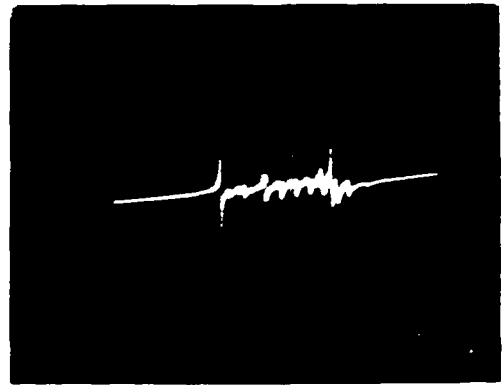


Figure 11: Spectrum of the third overtone of the blank of Figure 10 displaying the clean trace of the main response.

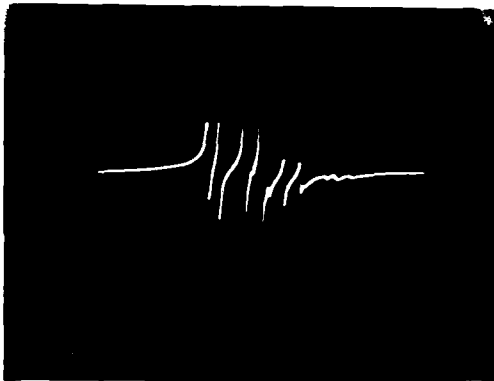


Figure 9: Spectrum of the third overtone of the same blank displaying the clean trace of the main response.

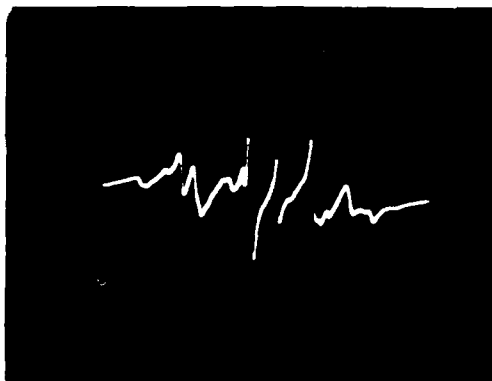


Figure 10: Spectrum displaying destruction of the fundamental main response by an interfering mode. This blank cannot be made to oscillate in a crystal impedance meter.

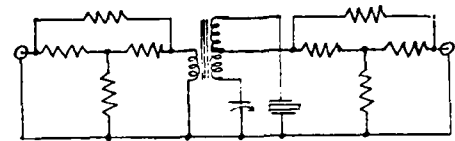


Figure 12: Bridge Network - Final Version. All resistors: 51 Ω 1/8 watt; Transformer; Mini Circuits Laboratory Model T4-1.



Figure 13: Photograph of Air-Gap Fixtures.

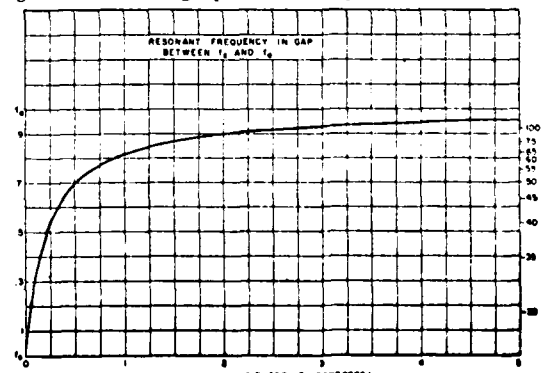


Figure 14: Measured frequency f_{r1} Interpolated between f_{s1} and f_{a1} , cf. Figure 4.

A NEW QUALITY EVALUATION METHOD OF RAW QUARTZ BY MEASURING
THE Q-VALUE OF Y-BAR RESONATOR

Hitohiro Fukuyo, Nobunori Oura and Fumiyasu Shishido
Tokyo Institute of Technology
4259 Nagatsuta, Midoriku, Yokohama, Japan 227

Summary

Two measurement methods have been adopted for the evaluation of raw quartz for resonator use, one of which is based on the measurement of absorption rate of the infrared ray of the specified wavelength and the other is based on the Q-value measurement of a 5th overtone 5 MHz planoconvex AT-cut plate with the specified shape and angle of cut. The former method is simple and convenient but is of the indirect presumption of elastic vibration loss of the quartz under test through the absorption rate. In the latter method, there must be a requisite that the measurement loss does not include any loss except elastic vibration loss in the quartz. However, this requisite seems not to be satisfied because of the complex structure of the resonator. As a result, resonators fabricated by different makers using the identical raw quartz show different measurement values from maker to maker. Therefore some improvements are required for a reliable measurement of intrinsic loss.

Our method is also based on the measurement of the Q-value of a resonator made from the raw quartz under test, however, to meet the requisite, a slim -18.5° cut Y-bar resonator was used. We measured the Q-value of the resonator mounted in the specially designed holder supporting the nodal line of motion placed on a very narrow metal pillow, and examined experimentally influences on the Q-value of all extrinsic losses due to the coupling with spurious modes, the lapped and polished layer, the support, the exciting electrodes, the acoustic radiation to peripheral gas and the contamination on surfaces. And we made clear the necessary conditions under which all extrinsic losses can be ignored. We confirmed that the Q-value measured under the above-mentioned conditions indicated only the intrinsic vibration loss in the quartz. According to this method, the resonators manufactured by different makers using the identical quartz give approximately the same results. The elastic vibration loss of the quartz cultured in different processes can also be discriminated through the Q-value. Finally some difficulty experienced in holding stably the resonator on the above-mentioned narrow pillow was eliminated by applying a new holder in which the resonator was supported at two nodal lines of the 2nd overtone mode by specially equipped two knife edges.

Introduction

In this paper all losses of crystal resonator having the delicate influences upon the Q-value are classified first and some explanations are described about the idea that the quality of raw quartz for resonator use must be evaluated reliably by measuring the Q-value of a test resonator made from that quartz if the necessary conditions to ignore the other losses except the elastic vibration loss in quartz crystal can be found experimentally and kept very carefully.

Then the details of a slim bar crystal resonator for this purpose and its special holder to be able to keep the conditions described are explained. After that, each characteristic of the losses except the inherent loss in quartz crystal is studied and the

necessary conditions to be able to ignore such losses practically are found mainly by experiments.

The measured Q-values of the test resonators made from some raw quartz cultured in different processes are shown and referring to the good correlation between their Q-values and culturing processes, the quality of quartz for resonator use is evaluated reliably by the proposed method. At the end of this paper an improvement of this method is described.

Losses of Crystal Resonator

As ωL_1 of a given crystal resonator is constant, the Q-value depends on its R_1 resulted from total losses of the resonator, where R_1 is the resistance of motional arm in the equivalent circuit of crystal resonator.

R_1 can be divided into various components as follows.

$$R_1 = R_m + R_g + R_w + R_h + R_e + R_r + R_c,$$

where R_m is a component resulted from the elastic vibration loss in quartz crystal,
 R_g is a component resulted from the loss caused by the coupling with spurious vibration,
 R_w is a component resulted from the elastic vibration loss in the lapped and polished layer of resonator,
 R_h is a component resulted from the elastic vibration loss due to the holding of resonator,
 R_e is a component resulted from the elastic vibration loss in the plated metal electrodes,
 R_r is a component resulted from the elastic vibration loss radiated into peripheral gas,
 R_c is a component resulted from the loss caused by the contamination left on the surfaces of resonator.

As the Q-value also depends on these resistance components, it is necessary that R_g, R_w, R_h, R_e, R_r and R_c should be so small as to be ignored comparing with R_m in order to evaluate the quality of raw quartz for resonator use by measuring the Q-value of a test resonator made from that raw quartz. In other words, total R_1 must be only R_m effectively.

Test Crystal Resonator

A -18.5° cut slim Y-bar crystal resonator with rectangular cross section shown in Fig.1 is used as a test resonator. The vibrational mode of this resonator is well known and the nodal line of the fundamental longitudinal vibration exists at the center of the length and lies exactly perpendicular to the direction of the length.¹

For convenience of fabrication of a test resonator from the quartz crystal, a resonator with geometry of $x_0 \pm 2.3$ mm and $y_0 \pm 25$ mm is chosen and applying $z_0/y_0 = 0.1$ as an adequate dimension's ratio without any spurious vibration in a mode chart, $z_0 \pm 2.5$ mm is also determined. The resonance frequency of this resonator is about 100 kHz.

Special Holder for Test Resonator

A gap type holder for a test resonator is prepared and its schematic construction is shown in Fig. 2.

The nodal line of the resonator is set exactly upon a narrow metal pillow equipped at the center of the lower electrode. Figure 3 shows this special lower electrode with a very narrow pillow.

This holder for the new evaluating method can also be used to investigate experimentally the complicated influences which the above-mentioned various losses have upon the Q-value of a test resonator.

Characteristics of Various Losses except the Loss in Raw Quartz and Necessary Conditions to ignore such Extrinsic Losses

(1) Loss caused by Coupling with Spurious Vibration

As the Q-value of the vibration coupled with other mode decreases as the increase of its coupling effect, the existence of such a coupled spurious vibrational mode must be checked carefully.

As the shape and dimensions of a test resonator for this method are chosen adequately from a mode chart, any coupled spurious vibration does not exist. Therefore the resistance component R_g resulted from the loss caused by the coupling effect can be ignored.

(2) Elastic Vibration Loss caused by Lapped and Polished Layer of Resonator²

As the elastic wave traveling in a resonator is attenuated on its reflecting surface with a lapped and polished layer, such a layer must be taken away from the surface of the resonator sufficiently to suppress the attenuation of the wave. Figure 4 shows the detail of structure of lapped layer enlarged by means of oblique cutting method. The depth of such a layer is found to be about 1/3 of the mean diameter of abrasive by the authors' fundamental investigation concerned with the worked layer of crystal plate.³

Figure 5 shows an example of lapping and polishing processes based on such an investigation.⁴ By using fine abrasive step by step, following this process, the lapped layer becomes very thin and the final polished layer is taken away by adequate etching. Then the desirable good surface on which any worked layer seems not to exist can be obtained.

In this study only four side planes of a test crystal resonator were lapped by some abrasives with different diameters at first, and measuring its Q-value one by one the relation between the Q-value and the mesh number of abrasive was examined.

Figure 6 shows its experimental result. The curve of the Q-value goes up and saturates gradually with the increase of mesh number and the Q-value of the resonator lapped with #3000 abrasive reaches about twice as large as that of #800.

After this experiment, the four side planes were lapped more with #5000 abrasive and polished. And then the both end planes of a resonator were also polished. In this case the Q-values more increased about twice and two and half times as large as that of #3000 abrasive shown in Fig. 6, respectively.

The above-mentioned experimental results show two important facts. One of them is that even the worked

layer with extremely thin depth comparing with the wavelength of 100 kHz elastic vibration has severe influence on the Q-value of the resonator and especially the lapped layer of four side planes which seems not to be important in reflecting the elastic wave has unexpectedly large influence on the Q-value of the resonator. The other is that the worked layer must be taken away sufficiently by lapping, polishing and adequately etching all 6 surfaces of the resonator following a systematic process in order to ignore the loss caused by worked layers in this method.

(3) Elastic Vibration Loss due to Holding of Resonator²

A resonator can be held stably at the nodal part of the vibration of interest. However, a little vibration energy seems to leak through the holding part practically and R_h component resulted from such energy loss reduces the Q-value of the resonator.

In this study a test resonator was put exactly at its nodal line on a metal pillow with various kinds of width such as 1, 0.5, 0.2, and 0.1 mm, and measuring the Q-value in each case the relation between the Q-value and the width of pillow was observed. Figure 7 shows this experimental result.

The Q curve goes up gradually with the reduction of pillow width and Q_{h0} which corresponds to the Q-value without any holding loss or in case of $R_h=0$ can be obtained as 28.3×10^4 . As the measured Q-value in 0.1 mm pillow width is 27.2×10^4 , and its deviation from Q_{h0} is small and only about -4%, the measured Q-value in 0.1 mm pillow width seems to be used as Q_{h0} .

In the experiment the difficulty to measure the Q-value, holding the resonator stably, was often experienced. In order to eliminate such difficulty and ignore the holding loss more sufficiently an improvement of the holder was tried later.

(4) Elastic Vibration Loss in Plated Metal Electrodes

The ohmic loss in the plated metal electrode with adequate thickness can be ignored practically. However, some elastic vibration loss can not be ignored generally and its loss reduces the Q-value of resonator.

In this method, the gap type metal plate electrodes were used regarding the fact that the elastic vibration loss in plated metal electrodes was related complicatedly with the thickness and structure of metal film. Therefore the above-mentioned loss needs not to be taken into account and R_e seems to be ignored practically.

(5) Elastic Vibration Loss radiated into Peripheral Gas²

This loss also reduces the Q-value of a resonator. In this study the relation between the Q-value and the peripheral air pressure was investigated setting a test resonator mounted in the holder in a vacuum bell jar. Figure 8 shows the experimental result.

Q_{r0} in vacuum or in case of $R_r=0$ can be obtained as 8.2×10^4 at the crossing point of the measured curve and the ordinate, and this value is very near to the measured Q-value of 7.94×10^4 in 3 Torr and its deviation rate is only about +3%. In other words the loss radiated into peripheral gas seems to be almost ignored practically in less than 3 Torr of gas pressure.

(6) Loss caused by Contamination left on Surfaces of Resonator²

This loss makes the measurement unstable and very often disturbs the measurement of the Q-value of a resonator.

In this method it is quite necessary to clean the surfaces of a test resonator carefully and dry it sufficiently to the utmost extent to ignore R_c before every experiment.

Quality Evaluation of Raw Quartz for Resonator Use by the Proposed New Method²

From some raw quartz crystals cultured in different processes of which any information was not given, four test resonators (A,B,C and D) were prepared and their Q-values were measured keeping very carefully the necessary conditions described before. The table shows these results.

They can be plainly classified into three groups - high group of 118.3×10^4 and 115.7×10^4 , medium group of 76.5×10^4 and low group of 43.0×10^4 - and these three kinds of the Q-value are found to correspond respectively to the different three culturing processes informed after the experiments.

This result seems to show that the quality evaluation of raw quartz for resonator use can be done reliably by this method if the necessary conditions described in this paper are kept sufficiently.

Improvement of the Holder⁵

In order to eliminate the practical difficulty encountered in making the measurement of the Q-value of a test resonator stably, a new holder was designed. This holder is shown in Fig.9.

In this holder the resonator is put stably at its two nodal lines of the 2nd overtone longitudinal vibration upon specially equipped two knife edges with the distance of half length of the resonator.

The maximum Q-value can be measured by adjusting precisely the mutual position between the resonator and two knife edges by means of mechanical stage. As this value corresponds to the Q-value in the case which holding loss can be ignored practically, the Q-value measured by using this holder seems to show more reliably the quality of raw quartz for resonator use. Figure 10 shows the characteristics of the Q-value vs air pressure measured by using the new holder. The interesting thing is that the high Q-value of more than 200×10^4 can be obtained in the improved method.

Conclusion

Various kinds of loss of a resonator which have the delicate influences upon the Q-value were analyzed first and the necessary conditions to be able to ignore practically the losses except the elastic vibration loss in quartz crystal were investigated.

A simple -18.5° cut slim Y-bar resonator to evaluate the quality of raw quartz and the specially designed gap type holder were explained. This resonator has not any spurious vibration which causes the coupling loss and the holder can hold the resonator under the necessary conditions which the losses except the loss in quartz crystal can be ignored practically.

The experimental result obtained from some raw quartz cultured in different processes showed that the quality evaluation of raw quartz could be done reliably by this proposed method if the necessary conditions

described in this paper were kept carefully.

As the principle of this method is very simple and clear and the measuring devices are also easy to be prepared, the quality evaluation of raw quartz by this method will be useful in the fields concerned.

Acknowledgements

The authors would like to thank the late Mr. T. Hashiguchi for his basic study on this problem at Tokyo Institute of Technology.

Thanks are also due to Dr. S. Nonaka who is now with Nippon Electric Co., Ltd. for his fundamental investigation on the lapped and polished layer of quartz at Tokyo Institute of Technology and the personnel concerned of Tokyo Communication Equipment Co., Ltd., Nihon Dempa Kogyo Co. and Citizen Watch Technical Laboratories for their cooperation throughout this work.

References

- 1) Hitohiro Fukuyo: Researches in Modes of Vibrations of Quartz Crystal Resonators by means of the "Probe Method", Bulletin of the Tokyo Institute of Technology, Series A, No.1, 1955.
- 2) Takeshi Hashiguchi: An Evaluation Method of Raw Quartz for Resonator Use, Master Thesis, Tokyo Institute of Technology, Feb. 1970.
- 3) Shunsuke Nonaka: Study on the Lapped and Polished Layer of Crystal Plates and Evaporated Electrodes for Crystal Resonator, Doctor Thesis, Tokyo Institute of Technology, March 1968.
- 4) Hitohiro Fukuyo: Crystal Resonator for Watch (3rd Lecture), Journal of the Horological Institute of Japan, No.73, March 1975.
- 5) Hitohiro Fukuyo, Nobunori Oura and Fumiyasu Shishido: The Quality Evaluation of Quartz by Measuring the Q-value of the Second Overtone of a Y-bar, Bulletin of the Tokyo Institute of Technology, No.131, pp 1-5, 1975.

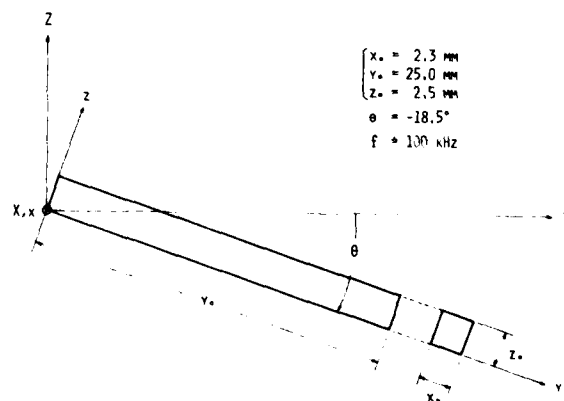


FIG. 1 TEST CRYSTAL RESONATOR

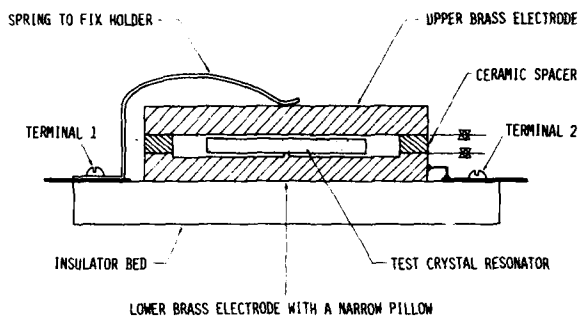


FIG. 2 AIR GAP TYPE HOLDER FOR TEST CRYSTAL RESONATOR

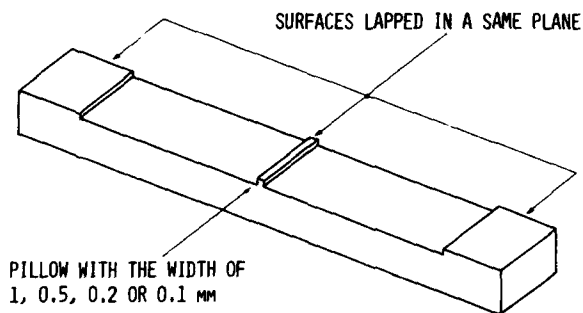


FIG. 3 LOWER BRASS ELECTRODE WITH A NARROW PILLOW TO HOLD CRYSTAL RESONATOR

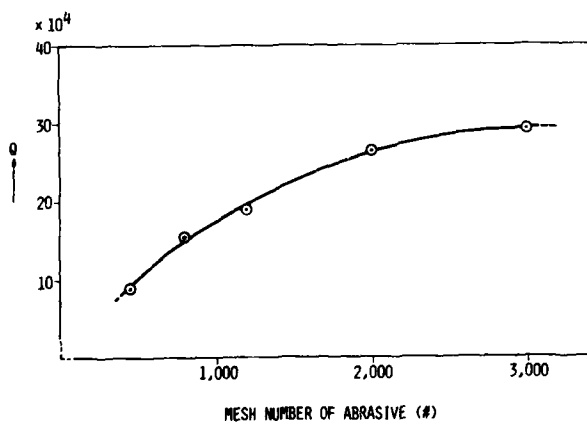


FIG. 6 RELATION BETWEEN Q AND FINENESS OF ABRASIVE

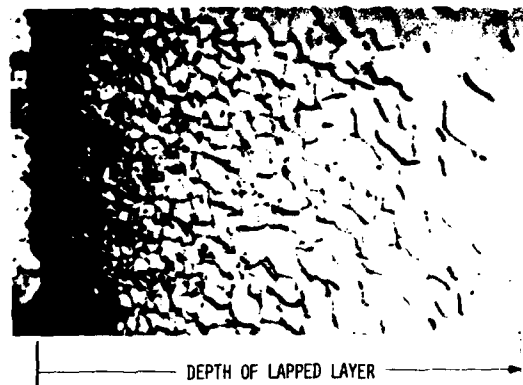


FIG. 4 LAPPED LAYER OF QUARTZ CRYSTAL OBSERVED BY MEANS OF OBLIQUE CUTTING METHOD

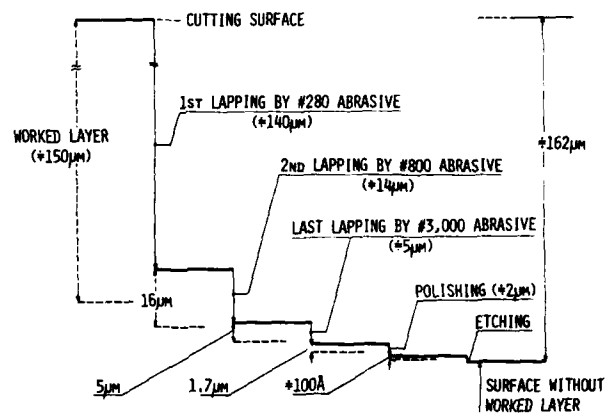


FIG. 5 AN EXAMPLE OF CUTTING, LAPPING, POLISHING AND ETCHING PROCESSES NOT TO LEAVE WORKED LAYER

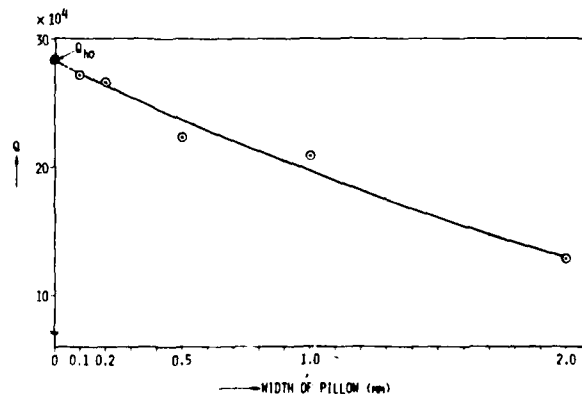


FIG. 7 RELATION BETWEEN Q AND WIDTH OF HOLDING PILLOW

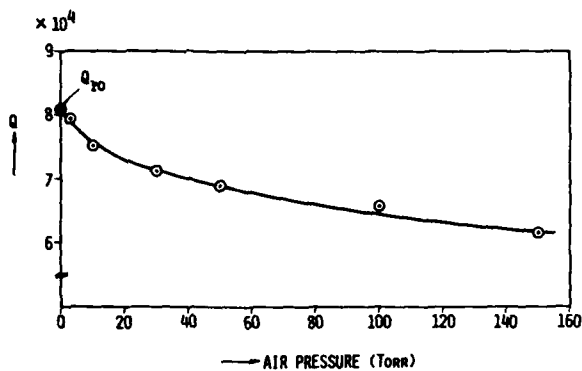


FIG. 8 RELATION BETWEEN Q AND PERIPHERAL AIR PRESSURE

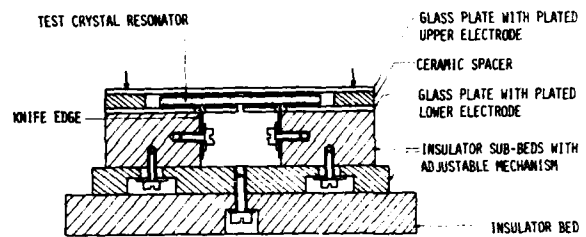


FIG. 9 IMPROVED AIR GAP TYPE HOLDER FOR 2ND OVERTONE MODE OF TEST CRYSTAL RESONATOR

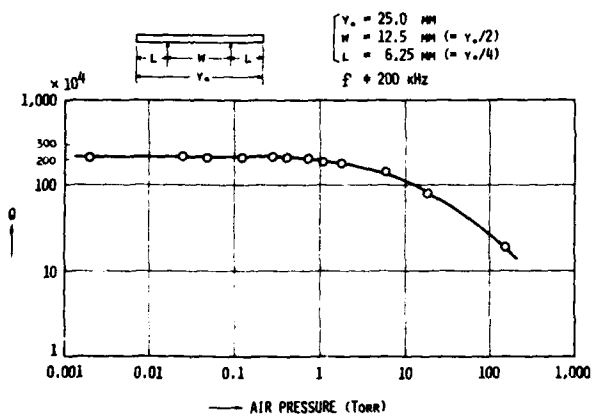


FIG. 10 RELATION BETWEEN Q IN IMPROVED HOLDER AND PERIPHERAL AIR PRESSURE

TEST RESONATOR	MEASURED Q-VALUE	CULTURING PROCESS
A	43.0×10^4	3RD CLASS
B	118.3×10^4	1ST CLASS
C	115.7×10^4	"
D	76.5×10^4	2ND CLASS

TABLE: MEASUREMENT OF THE Q-VALUES OF TEST RESONATORS MADE FROM SOME RAW QUARTZS CULTURED IN DIFFERENT PROCESSES

ACOUSTIC BULK WAVE RESONATORS AND FILTERS OPERATING
IN THE FUNDAMENTAL MODE AT FREQUENCIES GREATER THAN
100 MHz

Marc Berté

Research Center THOMSON-CSF

Domaine de Corbeville 91401 ORSAY - France

Summary

A localized ion beam milling has been used for increasing the fundamental frequency of acoustic bulk wave resonators. Monolithic filters in lithium niobate ($Y + 37^\circ$) and lithium tantalate (X) operating respectively at 220 MHz and 230 MHz have been realized. The relative bandwidth was equal to 1% and the insertion loss was under 5 dB. An AT cut quartz resonator working in the fundamental mode at 270 MHz with a Q value of 14,000 has also been obtained. In this case the vibrating membrane was 6 μm thick with a center portion of diameter 2.5 mm.

Key words Resonator, Filter, High Frequency, Quartz, Lithium tantalate, Lithium niobate, Ion beam milling.

Introduction

The frequency of a piezoelectric resonator depends upon the thickness of the resonant membrane. Using conventional processing technology the minimum thickness that has been obtained with a reasonable yield is about 35 μm for quartz. This corresponds to a fundamental frequency of about 50 MHz for an AT cut. Other piezoelectric materials such as lithium niobate (LiNbO_3) or lithium tantalate have been studied for only the past few years. Their top limit frequencies are about 70 MHz. Piezoelectric resonators working at odd harmonic overtones are used but in this case the bandwidth of filters is reduced by the square of the overtone. Moreover the noise of crystal oscillators using overtone is increased. Then new technologies have been investigated for increasing the fundamental frequency of piezoelectric resonators. Ion beam milling has been proposed by several authors for realizing a very thin membrane^{1,2} but only a preliminary result on a high frequency lithium niobate resonator has been published³. Experimental results on ion beam milled resonators and filters using lithium niobate, lithium tantalate and quartz are reported in this paper.

Ion beam milling technique

A thin piezoelectric disc is covered by a glass mask having a hole of 2 or 3 mm in diameter for limiting the milled zone. Both disc and mask, mounted on a rotating water-cooled support, are exposed to an argon ion beam. Thus only the center area of the disc is milled in order to obtain a thickness corresponding to the desired frequency. The very thin membrane is surrounded with a thicker rim. The handling of the resonator is relatively easy as a consequence of this structure. Generally the initial substrate obtained by a conventional polishing process is about 50 μm thick and 6 mm in diameter. The two faces are parallel and have a very good surface state.

The parallelism and polish of surface must be preserved during the ion milling because these features are very important for fabricating resonators

with a single resonance mode and a high Q value. Surface state preservation is dependent upon the angle between the ion beam and the milled surface. Optimum incidence angle of the ion beam on the surface of the piezoelectric materials was determined to be 35° . The angle between the rotation axis and the ion beam direction can be adjusted from 0 to 90° as shown in figure 1. Taking the argon ion energy as 1.8 KeV and the current density of 0.75 mA, the milling rate is about 3 $\mu\text{m}/\text{hour}$ for the materials used. The interference fringes were observed for controlling the parallelism of the membrane after milling. It appears that the parallelism of the crystal was not perturbed by ion milling.

Therefore very thin membranes with good physical properties can be obtained by ion milling techniques.

Design and fabrication

Energy trapping is a fundamental concept in the design of piezoelectric resonators. When energy trapping is maximum the resonator operates in a single mode, and no spurious resonances appear. The insertion loss at the resonance frequency is minimum and the Q value is maximum. The best energy trapping is obtained by adjusting the shape of metallic electrodes for lithium niobate ($Y + 37^\circ$ rotated cut) and lithium tantalate (X cut) because the electromechanical coupling coefficient is relatively high ($\approx 45\%$) and the influence of mass-loading is relatively small. For lithium niobate ($Y + 37^\circ$ rotated cut) electrodes dimensions must be about 2.3 times greater than the resonator thickness in order to achieve good trapping conditions in the fundamental mode³. For lithium tantalate (X cut) the best ratio between electrode dimensions and the thickness of the membrane is about 6.5 in the direction of vibration, called "thickness shear direction" (TS) and about 4. in the thickness twist direction (TT) perpendicular to TS^{4,5}. For quartz (AT cut) the electromechanical coefficient is small ($\approx 8\%$) and mass-loading becomes preponderant. Maximum energy trapping can be obtained by adjusting the electrode thickness. Moreover, the diameter of the electrodes must be about 20 times the membrane thickness^{6,7,8}. Monolithic crystal filters can be realized using pairs of electrodes deposited on the two faces of a thin membrane. The spacing between electrodes controls the coupling between both resonators. Usually this inter-electrodes spacing is slightly less than the crystal thickness. Metallic electrodes have a very little surface area and must be deposited with great accuracy. They are made by sputtering aluminium through metallic masks on the two faces of the membrane. Then resonators or filters are mounted in standard crystal packages and their frequency response is measured.

Experimental results

Lithium niobate, lithium tantalate and quartz have been used as piezoelectric materials.

Lithium niobate (LiNbO₃)

A monolithic crystal filter made of two coupled resonators has been fabricated on a LiNbO₃ (Y + 37°) disc, 6 mm in diameter and 50 μm thick. Ion beam milling time was 12 hours for obtaining a 14.6 μm thick membrane. Two square-electrodes with 30 μm sides were deposited on the top face. The spacing between them was about 13 μm. The ground electrode was deposited on the opposite face. This filter has been matched to 50 Ω and tested. The central frequency is 221.8 MHz and the relative bandwidth is 0.9%. Total insertion loss is 4 dB and minimum rejection is more than 25 dB (see figure 2).

A central frequency equal to 378 MHz has been reached with an 8.6 μm thick membrane. Lithium niobate is a piezoelectric material with a very high electromechanical coupling coefficient but unfortunately the frequency temperature stability of the Y + 37° cut is close to - 75 ppm/°C and such filters can be used only in thermostated devices.

For obtaining resonators and filters having a good temperature stability, discs of lithium tantalate (X) and quartz (AT) were ion-milled

Lithium tantalate (LiTaO₃)

A monolithic lithium tantalate filter operating at 230 MHz has been realized. The initial crystal were 6 mm in diameter and 50 μm thick. After milling, the membrane is 8.6 μm thick and its diameter is about 2 mm. Electrodes of 40 x 23 μm have been obtained using aluminium sputtering. They are separated by a width equal to 7 μm. The center frequency is 231.51 MHz and the 3 dB bandwidth is 4.28 MHz. This corresponds to a 1.85% relative bandwidth. The insertion loss is close to 5 dB after impedance matching to 50 Ω and the minimum rejection is 30 dB as shown in figure 3. The LiTaO₃ (X) resonance frequency variation versus temperature is parabolic with a turn-over temperature around 0°C and a second order coefficient equal to $52 \times 10^{-9}/^{\circ}\text{C}^2$.

Quartz (SiO₂)

At cut quartz resonators working at very high fundamental frequency were also studied and tested. A resonator working at 271.224 MHz has been realized. The initial substrate was a disc 6 mm in diameter. The thickness was 55 μm. It was milled for 15.45 hours and the final membrane was 6.1 μm thick. The sizes of aluminium electrodes deposited on the two faces of the membrane were adjusted by chemical etching for obtaining a single mode of resonance. The frequency response of this resonator does not show any spurious resonances. (See figure 4). The antiresonance frequency is 272.732 MHz and Q value is about 14,000. The effective electromechanical coupling coefficient is 7.3%. A schematic view of the dimensions of the resonator is given in the figure 5.

An AT quartz resonator has been studied for obtaining a single resonance mode at the third harmonic. The thin membrane is 16.8 μm thick. The third harmonic resonance frequency is 296.58 MHz and the antiresonance frequency is 296.64 MHz. Insertion loss of 13 dB and Q value of 18,000 have been measured. The frequency response does not show any spurious resonances. (See figure 6).

Conclusion

These preliminary results show the feasibility of very thin membranes (6 μm thick and 2.5 mm wide) using localized ion milling. These membranes appear to be rugged enough and can be easily handled and mounted. Thinner membranes can be probably obtained. Metallic electrodes have been deposited in correct position on the vibrating membrane. The parallelism of the membrane, and the quality of the surface state are good enough to obtain a clear response of the resonator and a useful Q value. Using this new technique a fundamental frequency of around 300 MHz has been reached for AT cut quartz resonator. Good characteristics have also been obtained for an AT cut quartz resonator operating at 300 MHz at the third harmonic.

Fundamental frequency of quartz resonators has been increased five times over the frequency of conventional resonators by means of a localized ion beam milling. Moreover very high frequency range can be reached using the third harmonic. The application of these high frequency components are very important because when the fundamental mode is used the bandwidth of filters is much larger than that of devices operating at odd harmonic. Voltage controlled oscillators with relatively wide frequency shifts and filters for telecommunications can be implemented advantageously with these new acoustic components.

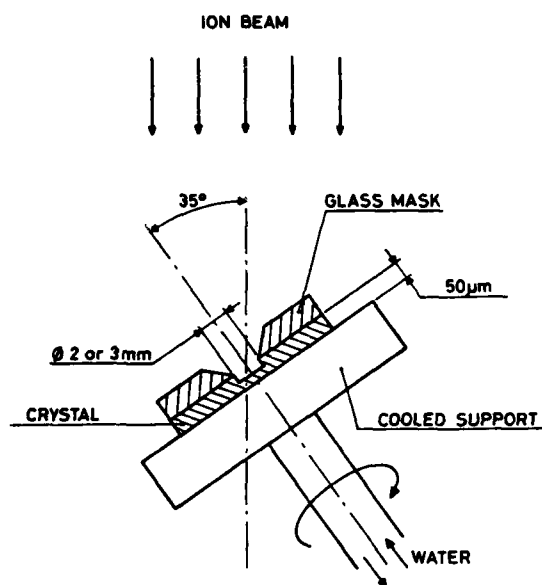
Acknowledgements

The author would like to acknowledge helpful discussions with P. Hartemann and the technical assistance of A. Soufaché, and H. Facchetti.

This work has been supported by the "Direction des Recherches et Moyens d'Essais" France.

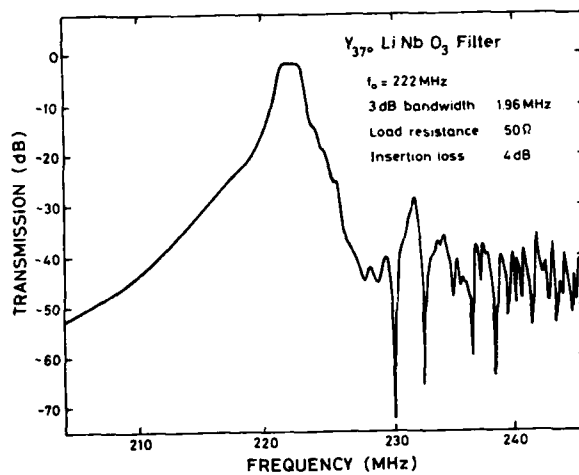
References

- 1 - R.N. Castellano and J.L. Hokanson "A Survey of Ion Beam Milling Techniques for Piezoelectric Device Fabrication" Proc. 29th Annual Symposium on frequency control, pp. 128 - 134, 1975.
- 2 - G.K. Guttwein, A.D. Ballato and T.J. Lukaszek. "United States Patent 3. 694,677 - 26 Sept. 1972
- 3 - G.C. Coussot " Bulk Wave Resonator for the 100-300 MHz Frequency range". Proc. IEEE, pp.590-591, 1974.
- 4 - K. Sawamoto "Energy Trapping in a Lithium Tantalate X cut Resonator" Proc. 25th Annual Symposium on Frequency Control. pp 246-250, 1971.
- 5 - G. Volluet and M. Savy " Etude d'Application Electronique du Tantalate de Lithium" Rapport DRME N° 36 803 195.
- 6 - W. Shockley, D.R. Curran and D.J. Koneval "Trapped Energy Modes in Quartz Filter Crystals" J. acoust. soc. Am, 41 pp. 981-983, 1967.
- 7 - W.D. Beaver. Theory and Design Principles of the Monolithic Crystal Filter. Ph. D. Dissertation Lehigh University, Bethlehem, Pa, 1967.
- 8 - G.E. Roberts - The Design of Coupled-Resonators AT-Cut Quartz Crystals for Operation on the Third Thickness-Shear Overtone - Proc. of the IEEE pp.1527-1528, Oct.1975

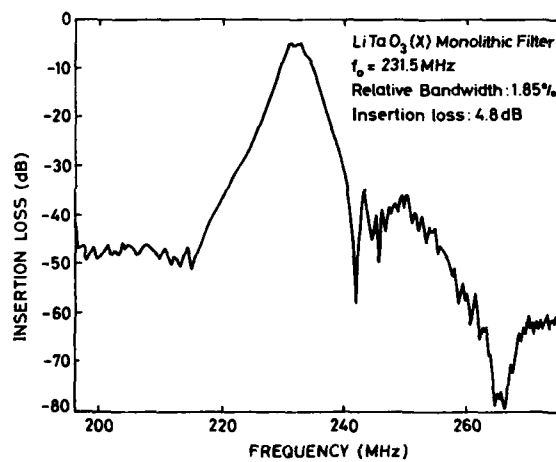


SCHEMATIC ION MILLING SUPPORT

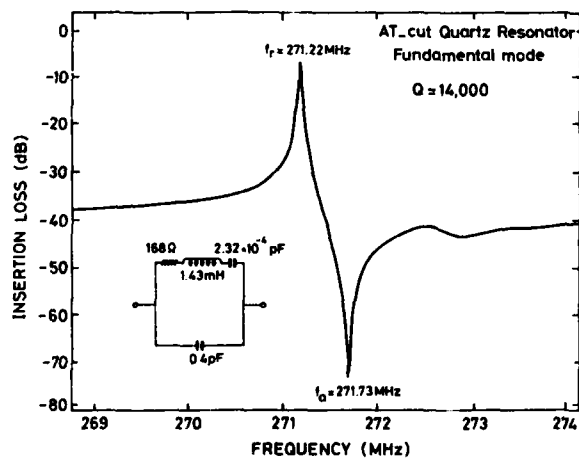
1 - Schematic of the ion milling support



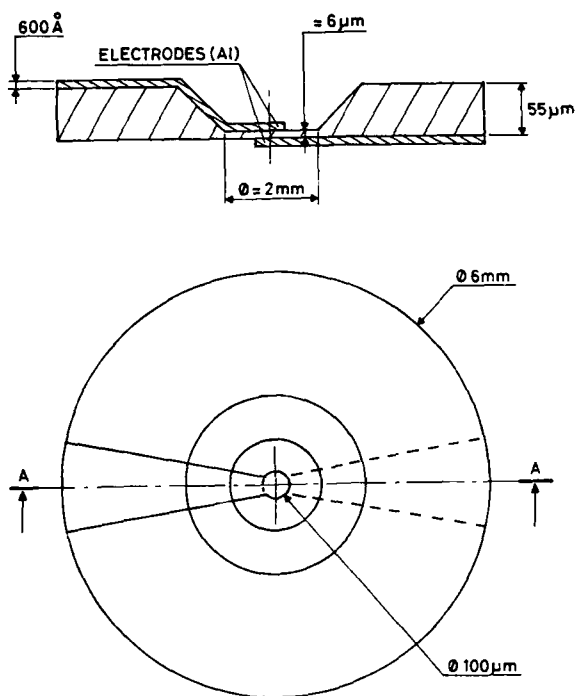
2 - Frequency response of a Y + 37° LiNbO₃ filter



3 - Frequency response of a X LiTaO₃ filter

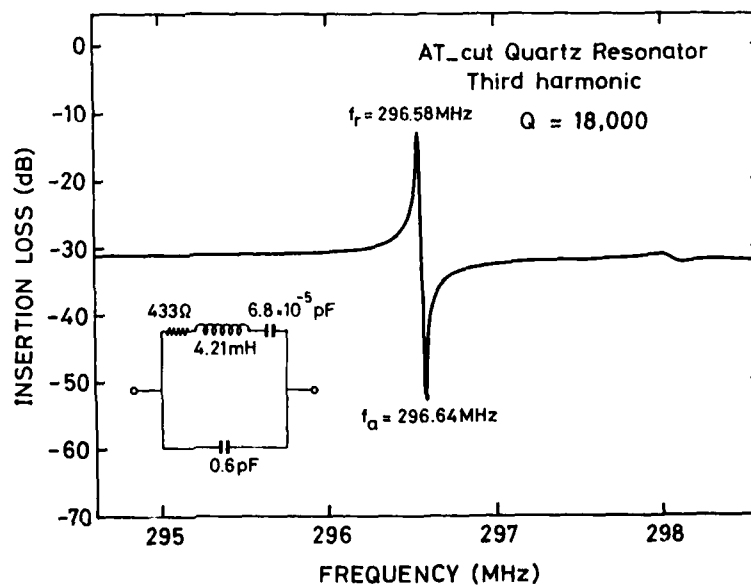


4 - Insertion loss of an AT Quartz resonator operating at the fundamental mode.



QUARTZ RESONATOR

5 - Schematic of the 271 MHz AT Quartz resonator



6 - Frequency response of an AT Quartz resonator operating on the 3rd harmonic overtone.

THE RELATIONSHIP BETWEEN QUARTZ SURFACE MORPHOLOGY AND THE Q OF HIGH FREQUENCY RESONATORS

R. N. Castellano, T. R. Mecker and R. C. Sundahl
Bell Telephone Laboratories, Incorporated
Allentown, Pennsylvania 18103

and

J. C. Jacobs
Western Electric Company
North Andover, Massachusetts 01845

Summary

Surface inhomogeneities caused by mechanical lapping contribute to surface stresses and energy dissipation and result in noise, rapid aging, drive sensitivity, and frequency versus temperature irregularities. As a result, several authors have stressed the need of utilizing polished quartz plates in order to achieve a Q limited only by the quartz itself.

An alternative approach evaluated in this study is to utilize a 3 μ m lapped and deep-etched plate in the fabrication of high frequency resonators. Sufficient material can be removed by the etching to essentially eliminate the damaged layer caused by the lapping operation.

The resonator Q reaches a maximum value at $\Delta f/f^2 = 1.6$ upon chemical etching and then decreases with further etching. At this point, there is little loss in Q or device performance when compared to resonators made from polished plates. The change in Q with increasing $\Delta f/f^2$ is proportional to the resonator frequency, whereas the value of $\Delta f/f^2 \text{ MAX}$ is independent of resonator frequency.

The overall surface of these plates appears to become smoother with increasing chemical etching. However, close examination reveals that the depths of the surface striations increases as the amount of material removed increases. At a removal of $\Delta f/f^2 = 0.73$, the depth of the striations is approximately 0.3 μ m and the width is 5 μ m. As more material is removed during etching, the striations become broader and extend deeper into the plate.

The Q degradation phenomenon is shown to be correlated with the depth (d), and width, (w), of surface striations initiated by mechanical lapping and enhanced by chemical etching. A Rayleigh scattering model is proposed to explain the degradation in resonator Q by the elastic scattering of an acoustic wave by these striations. This model predicts that the degradation in Q should follow the form:

$$Q^{-1} = K (d^2 w^2 f^4)$$

where K is a constant. This equation can be used to predict the loss in Q of resonators made from lapped and etched plates as a function of frequency and extent of chemical etching.

Introduction

It is well known that surface imperfections can adversely affect the performance of piezoelectric resonators¹⁻³. Surface inhomogeneities caused by mechanical lapping contribute to surface stresses and energy dissipation and result in noise, rapid aging, drive sensitivity, and frequency versus temperature irregularities. For high frequency resonators, this damaged layer is normally removed by polishing followed by light chemical etching. As a result, several authors^{1,2} have stressed the need of utilizing polished quartz plates in order to achieve a Q (quality factor) limited only by the quartz itself. In addition, a polished plate has a lower surface area than a lapped plate, which could result in less adsorption of residual contaminants and a lower aging rate. However, a polished plate is subject to minor scratches during the high friction of the mechanical polishing operation. Also, the adherence of gold electrodes is poor so that careful handling is necessary. Finally, the polishing operation contributes to the cost of the plate.

An alternative technique is to utilize a lapped and deep-etched plate in the fabrication of high frequency resonators. Sufficient material can be removed from each major surface to essentially eliminate the damaged layer caused by the lapping operation. The deep-etching reduces the surface area and the amount of adsorbed contaminants. In addition, savings in the cost of the plate could be realized by the elimination of the polishing operation.

Miller³ has shown that the amount of chemical etching is an important parameter in the performance and Q of 8 MHz fundamental mode monolithic crystal filters. The Q reaches a maximum at a removal of 50 KHz and remains level upon subsequent etching to 1000 KHz removal. Polishing increases the Q only 10%.

This memorandum describes an investigation into the relationship between surface morphology and performance of high frequency resonators. An optimum etch condition has been determined. In addition, the results of this and other investigations are interpreted in terms of a Rayleigh scattering model, suggesting the existence of an optimum etch condition which varies with resonant frequency.

Q Degradation

Natural AT-cut quartz plates were lapped with a 3 μ m or 5 μ m abrasive and etched in a

saturated ammonium bifluoride solution at 50°C. Figure 1 illustrates the change in Q with increasing chemical etching of 28 MHz, 5th overtone resonators. The resonator Q reaches a maximum value upon chemical etching and then decreases with further etching. The initial increase is believed due to the removal of surface damage caused by lapping. This maximum occurs at a removal of 1.4 μm from the lapped surface corresponding to a $\Delta f/f^2_{\text{MAX}}$ of 1.6.

At this point, there is little loss in Q or device performance when compared to resonators made from polished plates. The change in Q is less pronounced as the resonator frequency decreases, whereas $\Delta f/f^2_{\text{MAX}}$ is independent of resonator frequency.

Surface Morphology

The overall surface of these plates appears to become smoother with increasing chemical etching when evaluated with an SEM (Figures 2 and 3). This seems to contradict the data in Figure 1, since a polished plate, free of surface roughness, produces a device with a high Q (3.65×10^5). However, close examination reveals that the depths of the surface striations caused by the anisotropy of chemical etching increases as the amount of material removed increases. This is supported in Figures 2 and 3 where profilometer traces of these plates show the actual surface roughness. At a removal of 118 KHz ($\Delta f/f^2 = 0.7$), the depth of the crevices extends approximately 0.3 μm into the plate and has a width of 5 μm . As more material is removed during etching, the crevices become broader and extend deeper into the plate. At a removal of 3300 KHz ($\Delta f/f^2 = 20.8$) the depth of these crevices is 1.7 μm and the width is approximately 60 μm .

There is little difference in the profilometer traces between the unetched plate and the plate etched 118 KHz. However, quartz surfaces investigated by reflection high energy diffraction (RHEED)⁵ indicate that the surface of the unetched plate is mostly amorphous whereas the surface of the plate etched 118 KHz is entirely crystalline (the RHEED pattern consists of very sharp, well developed Kikuchi lines). Estimates of typical crevice widths (w) and depths (d) are listed in Table I, together with appropriate frequency, Q values, and etch conditions.

It is believed that the decrease in resonator Q with increased chemical etching is due to the effect of the surface morphology of the quartz plate.

Rayleigh Model

The experimental observations in Figure 4 can be interpreted in terms of a Rayleigh scattering model. Rayleigh⁴ has shown that the elastic scattering of an acoustic wave by a single particle is given by:

$$\frac{S.A.}{I.A.} = \frac{\pi T}{R\lambda^2} (1 + \cos \theta) \quad (1)$$

where S.A. is the scattered wave amplitude,

I.A. is the incident wave amplitude, R is the distance of the particle from the point of observation, T is the volume of the particle, λ is the wavelength, and θ is the angle between the direction of observation and the direction of the incident wave. The model is valid for cases in which $\lambda > T^{1/3}$.

This model has been applied to the scattering of acoustic waves by grains in polycrystalline metals⁶. In this present case, the particles are the surface discontinuities produced by the mechanical lapping and deep-etching.

The total energy scattered (S.E.) from a single particle is proportional to the square of the scattered amplitude integrated over the angles of scattering so that for N particles:

$$\text{Total S.E.} = (I.A.)^2 \frac{K}{\lambda^4} \sum_{i=0}^N T_i^2 \quad (2)$$

where K is a constant. Since the volume of a particle or crevice (T_i) is approximately proportional to the square of the cross sectional width (w) times the depth (d), then from the summation in equation (2):

$$\sum_{i=0}^N T_i^2 \propto N(w^2d)^2 \quad (3)$$

However, the total area over the plate is

$$A_{\text{tot}} = N w^2 \quad (4)$$

so that equation (3) becomes:

$$\sum_{i=0}^N T_i^2 \propto A_{\text{tot}} w^2 d^2 \quad (5)$$

Since $A_{\text{tot}} (I.A.)^2$ is proportional to the total incident energy from equation (2),

$$\frac{\text{Total S.E.}}{\text{Total I.E.}} = K' \frac{w^2 d^2}{\lambda^4} \quad (6)$$

The quality factor or Q of a resonator is given by

$$Q = 2\pi \frac{(\text{energy stored})}{(\text{energy dissipated per cycle})} \quad (7)$$

Substituting into equation 6 gives

$$Q^{-1} = K'' \frac{w^2 d^2}{\lambda^4} = K''' w^2 d^2 f^4 \quad (8)$$

This model predicts that as the width and depth of the crevices increases, the Q is degraded. It also predicts a decrease in Q for a given etch condition (constant w,d) as f increases, which is in agreement with Warner¹.

The measured Q of a device is affected by several parameters: the intrinsic Q of the quartz, the energy losses due to mounting, and the energy losses caused by surface

stresses and imperfections.

$$Q_{\text{measured}}^{-1} = Q_{\text{bulk}}^{-1} + Q_{\text{mounting}}^{-1} + Q_{\text{surface}}^{-1} \quad (9)$$

(lapped)

For a polished plate with a highly crystalline surface Q_{surface}^{-1} is negligible so that

$$Q_{\text{measured}}^{-1} = Q_{\text{bulk}}^{-1} + Q_{\text{mounting}}^{-1} \quad (10)$$

(polished)

The effect of the surface morphology on the Q of the resonator is thus

$$Q_{\text{surface}}^{-1} = Q_{\text{measured}}^{-1} - Q_{\text{measured}}^{-1} \quad (11)$$

(lapped) (polished)

So that equation (8) now becomes:

$$Q_{\text{surface}}^{-1} = K'' w^2 d^2 f^4 \quad (12)$$

Thus a plot of $\ln Q_{\text{surface}}^{-1}$ versus $\ln w^2 d^2 f^4$ should be linear with a slope of 1.0. Data in Table I is plotted in Figure 4 and agrees quite well with the above Rayleigh scattering model. The slope of this curve, plotted for several frequencies and device codes is 0.77 as calculated from a least square fit program. From the experimental data, equation (12) becomes:

$$Q_{\text{surface}}^{-1} = 5.57 \times 10^{-12} (d^2 w^2 f^4)^{0.77} \quad (13)$$

Energy loss due to the perturbation of the plate surfaces by lapping and etching is thus a function of the frequency and morphology of the quartz plate. A plot of dw versus $\Delta f/f^2$ is given in Figure 5. Values extracted from this curve can be inserted into equation (13) to estimate the loss of Q for a high frequency device.

Conclusion

It has been observed that the Q of high frequency resonators reaches a maximum value upon chemical etching and then decreases with further etching. This maximum occurs at a removal of $1.4 \mu\text{m}$ ($\Delta f/f^2 = 1.60$) from the $3 \mu\text{m}$ abrasive lapped surface. Little loss in Q or device performance is discernible when compared to resonators made from polished plates.

The decrease in Q upon subsequent chemical etching is due to the increase in the width and depth of crevices introduced to the plate by mechanical lapping and enhanced by chemical etching. A Rayleigh scattering model has been proposed to explain the degradation of Q by the elastic scattering of an acoustic wave by these crevices. Equation 13 can be used to predict the loss in Q of resonators made from lapped and deep-etched plates at various frequencies compared to those devices made from polished plates.

References

1. Warner, A. W., Bell System Technical Journal, XXXIX, 5, 1193-1217 (1960).

2. Smagin, A. G., Kristallografiya, 4, 6, 862-866 (1959).
3. Miller, A. J., Proc. of 24th Annual Symposium on Frequency Control, USAEC, 93-103 (1970).
4. Rayleigh, Theory of Sound, Vol. II, p. 152, (MacMillan, N. Y. 1929).
5. Vig, J. R., Cook, C. F., Schwidtal, K., Lebus, J. W., Hafner, E., Proc. of 28th Annual Symposium on Frequency Control, USAEC, 96-108 (1975).
6. Mason, W. P. and McSkimin, H. J., J. Acoust. Soc. Am., 19, 464-473 (1947).

EFFECT OF SURFACE REMOVAL BY CHEMICAL ETCHING ON Q OF 28 MHz RESONATORS

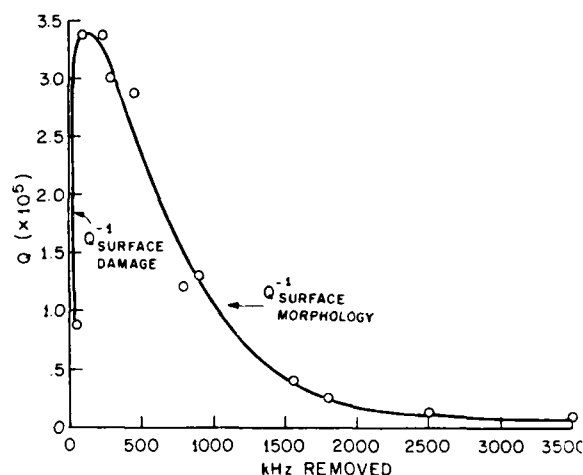
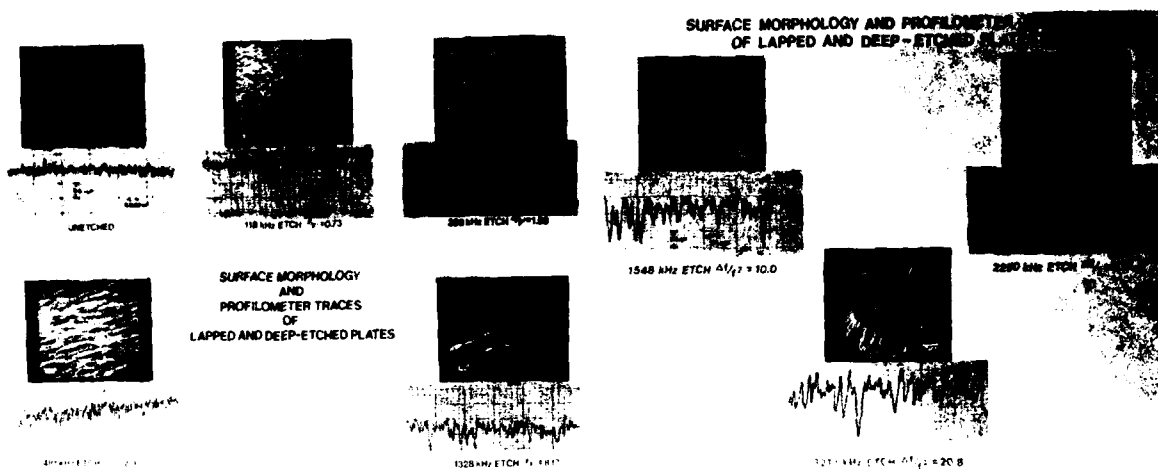


FIG. 1 Change in Q with increasing chemical etching in kHz of 28 MHz, 5th overtone resonators.



FIGS 2 and 3. Scanning electron micrographs and profilometer traces with different amounts of chemical etching.

EFFECT OF SURFACE MORPHOLOGY AND FREQUENCY ON Q OF HIGH FREQUENCY RESONATORS

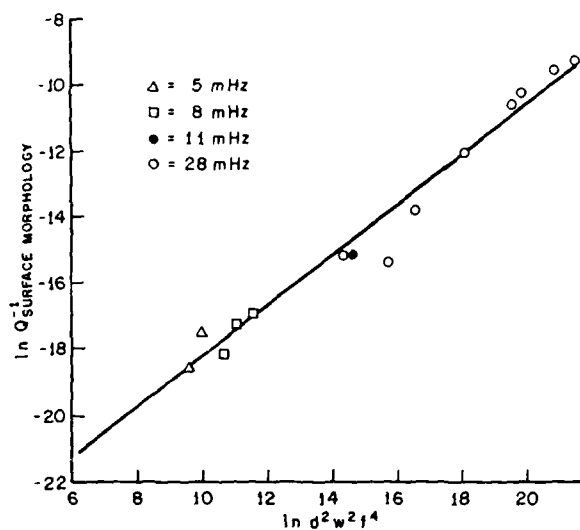


FIG. 4 Effect of surface morphology and frequency on Q_{surface} of various precision resonators. Natural logarithms are plotted because of the five decade spread of Q^{-1} values.

EFFECT OF CHEMICAL ETCHING ON SIZE OF CREVICES OF QUARTZ PLATES

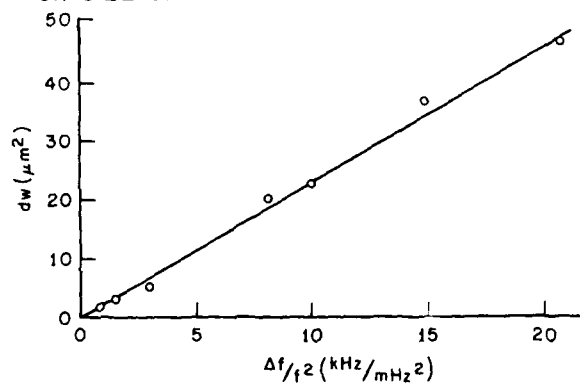


FIG. 5 Relationship between the size of the crevices and the amount of chemical etching.

TABLE I

Resonant Frequency (MHz)	Removal $\Delta f/f^2$	Q_{Measured}	Q_{Measured}^{-1}	Q_{Polished}	Q_{Polished}^{-1}	Q_{Surface}^{-1}	$(dw)^2$
4.950	1.70	2.35×10^6	4.26×10^{-7}	2.40×10^6	4.17×10^{-7}	8.53×10^{-9}	10.2
4.980	3.46	2.27×10^6	4.41×10^{-7}	2.40×10^6	4.17×10^{-7}	2.39×10^{-8}	36.0
8.216	1.71	1.57×10^6	6.37×10^{-7}	1.60×10^6	6.25×10^{-7}	1.19×10^{-8}	10.2
8.226	2.47	1.53×10^6	6.54×10^{-7}	1.60×10^6	6.25×10^{-7}	2.86×10^{-8}	20.2
8.235	3.11	1.48×10^6	6.76×10^{-7}	1.60×10^6	6.25×10^{-7}	5.07×10^{-8}	23.5
11.150	4.85	8.00×10^5	1.25×10^{-6}	1.00×10^6	1.00×10^{-6}	2.50×10^{-7}	180.0
28.567	0.86	3.12×10^5	3.21×10^{-6}	3.65×10^5	2.74×10^{-6}	4.65×10^{-7}	3.0
28.638	1.56	3.40×10^5	2.94×10^{-6}	3.65×10^5	2.74×10^{-6}	2.03×10^{-7}	10.2
28.231	2.96	2.70×10^5	3.70×10^{-6}	3.65×10^5	2.74×10^{-6}	9.61×10^{-7}	23.5
28.649	4.84	1.18×10^5	8.47×10^{-6}	3.65×10^5	2.74×10^{-6}	5.73×10^{-6}	180.0
28.925	8.03	3.90×10^4	2.56×10^{-5}	3.65×10^5	2.74×10^{-6}	2.29×10^{-5}	400.0
29.162	9.46	2.65×10^4	3.77×10^{-5}	3.65×10^5	2.74×10^{-6}	3.50×10^{-5}	506.0
29.883	13.62	1.40×10^4	7.14×10^{-5}	3.65×10^5	2.74×10^{-6}	6.88×10^{-5}	1296.0
30.856	18.71	1.12×10^4	8.93×10^{-5}	3.65×10^5	2.74×10^{-6}	8.68×10^{-5}	2070.0

CHEMICALLY POLISHED QUARTZ

John R. Vig, John W. LeBus and Raymond L. Filler
US Army Electronics Technology & Devices Laboratory (ECOM)
Fort Monmouth, NJ 07703

Summary

Etching in a saturated solution of ammonium bifluoride is shown to be capable of producing chemically polished AT-cut quartz surfaces over a broad range of conditions. The quality of chemical polish depends primarily on the surface finish prior to etching, the depth of etch and the quality of quartz used. The speed of polishing depends primarily on the temperature of the etching bath. In an 88°C etching bath, starting with 3 μ m lapped blanks, chemically polished blanks with a surface roughness of 0.1 μ m and a roughness angle of 1° can be produced in 15 minutes. Starting with a finer surface finish can produce a smoother chemically polished surface.

Chemically polished blanks are shown to be extremely strong. Fundamental mode 20 MHz resonators made with chemically polished natural quartz blanks showed no Q degradation with increasing depth of etch.

Of sixteen different varieties of cultured quartz from ten different growers, most were found not to be suitable for chemical polishing because of a large number of etch channels generated by the etching. The incidence of channels varied greatly from cultured quartz to cultured quartz. A vacuum swept cultured quartz was the only variety from which chemical polishing could produce blanks free of etch channels. Of a large number of 6.4mm diameter chemically polished natural quartz blanks, approximately 2/3 were free of etch channels.

Key Words. Etching, Polishing, Quartz, Quartz Resonators, Quartz Crystals, Etch Pits, Etch Channels, Dislocations, Cultured Quartz, Ammonium Bifluoride, Chemical Polishing, Lapping

Introduction

High precision and high frequency quartz resonators, particularly those for high shock applications, require quartz plates whose surfaces are free of imperfections, such as scratches and pits. The most common method of achieving such surfaces is mechanical polishing. The mechanical polishing processes are difficult to control (i.e., it is difficult to produce defect-free surfaces, at the correct frequency, with a high yield). Moreover, as has been known since the last century, even when the polished surfaces appear to be free of defects when examined at high magnification, the surfaces contain hidden defects^{1,2}. These defects can be revealed by X-ray diffraction topography³ and by etching subsequent to polishing.

Figure 1 is an SEM micrograph which shows the two most commonly observed defects revealed by etching polished quartz blanks: scratch marks and pits. As we shall discuss later, the pits often extend deep into the blanks.

The objective of the study to be described in this paper was to overcome the difficulties associated with mechanical polishing by finding an etching procedure capable of producing chemically polished quartz



Fig. 1. Polished surface after etching.
crystal surfaces.

Chemical Polishing

At the beginning of this study a literature search was conducted to determine what previous studies, if any, have been reported on the subject of polishing quartz chemically. No such reports were found. Perhaps this is due to the fact that it has been widely known that the etch rates of quartz vary greatly with crystallographic direction^{4,5}. The rate along the fastest etch direction, the Z direction, is over one hundred times the rate along the slowest etch direction, which is the X direction for right handed quartz. At first glance, therefore, achieving a chemically polished quartz surface may seem like a hopeless task.

It has, however, been shown that other highly anisotropic materials, such as silicon and germanium, can be polished chemically⁶. To understand why such anisotropic materials can be so polished, let us examine the etching mechanism.

Etching can be considered to be a five step process. The etchant must:

1. diffuse to the surface,
2. be adsorbed,
3. react chemically;

the reaction products must

4. be desorbed
- and 5. diffuse away from the surface.

The etching rate may be limited by any one of these steps. In chemical polishing, the rate controlling step is generally the diffusion to or from the surface. Diffusion control means that, in particular, the inherent

rate at which a reaction takes place at the surface is higher than the rate of diffusion; i.e., the etchant molecules at the surface react at a rate which is faster than the rate at which the concentration at the surface can be replenished by the diffusion of other etchant molecules. A depleted surface layer therefore exists, outside which the etchant concentration is uniform, but inside which the concentration decreases to near zero at the surface.

Under such conditions, the etching is principally determined not by the properties of the surface being etched, but by the diffusion. It is clear that if a surface initially consists of hills and valleys, as shown in Fig. 2., the probability of an etchant molecule diffusing to the top of a hill will be much greater than the probability of it diffusing to the bottom of a valley. The hills will therefore be etched faster than the valleys, and the surface will become increasingly smooth as the etching progresses.

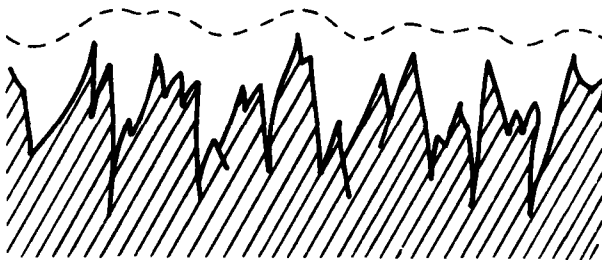


Fig. 2. A rough surface at the beginning of etching.

Eventually, the surface becomes so smooth that the depleted layer can have uniform thickness everywhere, e.g., as shown in Fig. 3. From that point, the surface is etched evenly everywhere, and the surface smoothness no longer improves with further etching. Chemically polished surfaces are therefore not perfectly flat but are microscopically undulating.

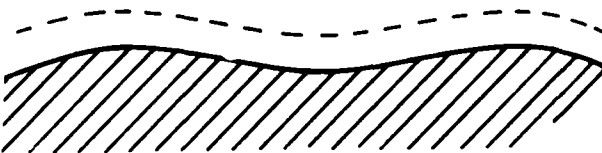


Fig. 3. Chemically polished surface.

Etching Experiments

Not enough information exists at the present for one to be able to predict the conditions under which a given etchant can chemically polish quartz. At the beginning of this study, an empirical evaluation of a number of the known etchants for quartz was planned. The evaluation began with ammonium bifluoride ($\text{NH}_4\text{F} \cdot \text{HF}$), simply because that had been the etchant used in our fabrication process, and was therefore the most readily available. As it turned out, we need not have looked any further.

Most of the experiments were performed on AT-cut plano-plano natural quartz blanks which had been final lapped with 3 μm MICROGRIT aluminum oxide abrasive⁷, then etched under various conditions in increments up to $\Delta f = 160f_0f_f$, where Δf is the change in frequency in kHz, and f_0 and f_f are the initial and final frequencies, respectively, in MHz. In the past, the depth of etch has usually been expressed in units of f^2 because the

depths of etch normally used in crystal fabrication have been small enough so that it did not matter if f meant f_0 or f_f . For the deep etches evaluated in these experiments the proper unit is f_0f_f . (Since $t_0 \propto \frac{1}{f_0}$ and

$$t_f \propto \frac{1}{f_f}, \Delta t \propto \frac{\Delta f}{f_0 f_f}.)$$

For AT-cut quartz, $160f_0f_f = 1.66 \mu\text{m}$ removed from the thickness.

Surface Evaluation

The etched blanks were first inspected with a light microscope to assure that no gross defects were present. The blanks were then examined with a scanning electron microscope (SEM), and with a Talysurf profile meter⁸.

The topography of deeply etched AT-cut quartz surfaces consist of hills and valleys which extend in a direction about 4° from the X axis. Therefore, in both the SEM and profile meter measurements, the topography appears smoothest when examined along a direction parallel to the hills and valleys, and appears roughest when examined perpendicular to the hills and valleys. Fig. 4 shows SEM micrographs of the same surface taken from the two different angles. These SEM micrographs, as well as all the others in this paper, were taken along an observation angle of 60° from the normal in order to provide good contrast and resolution. The micrograph on the left was taken looking along the hills and valleys; the micrograph on the right was taken looking across the hills and valleys. The scale is represented by an oval because of the oblique observation angle used.

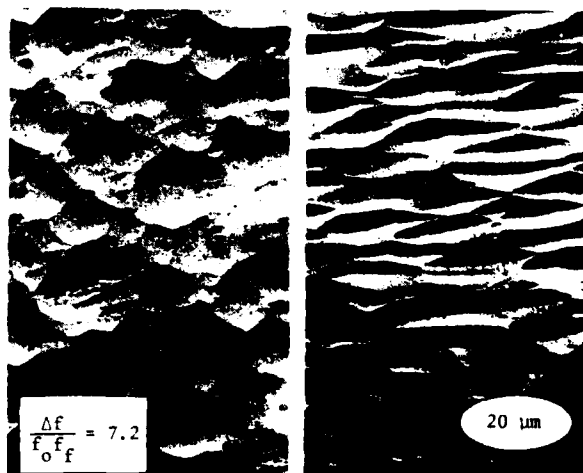


Fig. 4. SEM micrographs of the same surface from two different observation angles.

Similarly, Fig. 5 shows profile meter scans for the same surface along the two orthogonal directions. The profile meter is a stylus type instrument, similar to a phonograph. A very fine diamond needle is moved across the surface. The up and down displacements of the needle are converted to an electrical signal which is displayed on a chart. Since the depths of the valleys were generally small compared to the separations of the hills, the vertical scale in this example is magnified 200 times as compared to the horizontal scale. The surface appears roughest along the Z direction, i.e., the direction where the needle is moving perpendicular to the hills and valleys. It is, therefore, this direction on which all calculations of surface roughness in this paper are based.

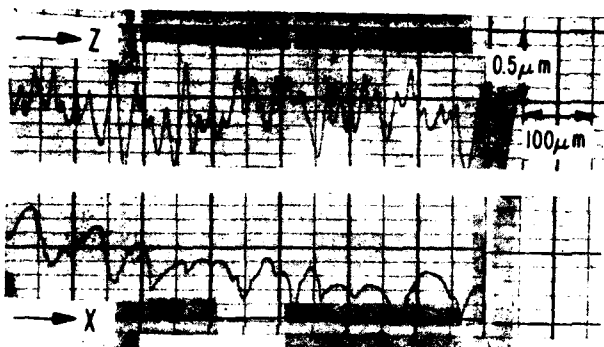


Fig. 5. Profile meter scans of the same surface along two orthogonal directions.

The Z direction profile meter scans were digitized⁹, and from this data the surface topography was characterized for each sample by two parameters: the surface roughness and the roughness angle. The surface roughness was derived by calculating the root mean square deviation from an imaginary, smooth "zero spatially averaged" line which is chosen so that the area of mountains above this line is equal to the area of valleys under this line. The roughness angle is the average angle which the sides of the hills make with the plane of the crystal blank.

Etching Apparatus

The etching experiments were performed in the double beaker arrangement shown in Fig. 6. A 1000 ml glass outer beaker contains water and a floating 400 ml Teflon beaker¹⁰, which in turn contains the saturated ammonium bifluoride solution. A temperature controlled stirring hot plate with a thermistor sensor¹¹ was used to control the temperature of the water surrounding the Teflon beaker. The temperature of the ammonium bifluoride solution could thus be controlled to about $\pm 1.5^\circ\text{C}$. A thick Teflon disc with a diameter slightly larger than the outer beaker was used as a cover to minimize evaporation from the beakers. The weight of this disc also served to push the inner beaker down to assure that the fluid level in the inner beaker was always about 3 cm below the water level in the outer beaker. A hole through the center of the disc permitted the agitation of crystals during etching.

Since the thermal conductivity of the Teflon beaker is low, to minimize the time required for the temperature of the inner bath to reach equilibrium with the outer bath, when the hot plate was first turned on, the inner bath was preheated in boiling water in a separate container. Even with this method it took well over an hour after the hot plate was turned on for the inner bath to stabilize at 85°C , for example. In production, it may be preferable to use direct heating of the solution by means of an alumina or Teflon coated immersion heater.

The crystals were held in a Teflon jig which was designed to assure that only point contacts existed between the jig and the crystals. For crystals with a diameter greater than 16 mm a commercially available jig¹² was found to be adequate. For the smaller crystals used throughout most of our experiments, a small "home made" Teflon jig was used, which fit on the same Teflon handle as the commercial device.

The crystals were agitated slowly in both directions by means of a constant speed electric motor, a (now obsolete) Hewlett-Packard Model 297A Sweep

Drive. During most of the experiments, the motor was set to rotate the etching jig through an angle of approximately 360° before reversing direction. The rate of agitation was about 5 cycles per minute.

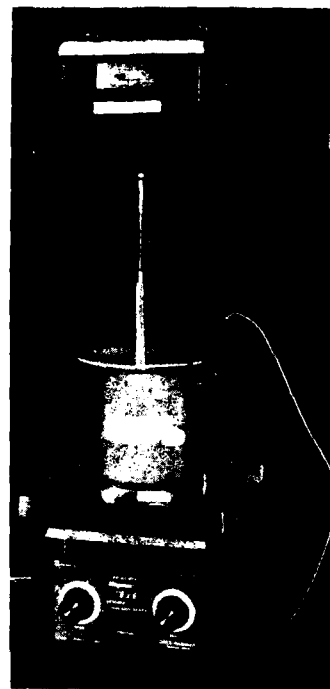


Fig. 6. Etching Apparatus

Etching Procedure

The etching procedure consists of the following steps:

1. A saturated solution of ammonium bifluoride is prepared in a Teflon container. The ammonium bifluoride ($\text{NH}_4\text{F}\cdot\text{HF}$) flakes¹³ are mixed with distilled water, and the solution heated to the desired temperature. The amount of $\text{NH}_4\text{F}\cdot\text{HF}$ used is such that after the solution reaches the equilibrium temperature, some undissolved flakes remain in the bottom of the container throughout the etching process. (The solubility of $\text{NH}_4\text{F}\cdot\text{HF}$ in water increases from 61 gms per 100 ml of solution at 60°C to 86 gms per 100 ml at 100°C ¹⁴.) The solution preparation and the etching are performed under a vented hood to prevent inhalation of the vapors from the etching bath.

2. The crystals are cleaned thoroughly. To assure that the surfaces are etched evenly, it is particularly important to remove all contaminants such as waxes and greases, which may be impervious to the etchant. The crystals used in these experiments were thoroughly degreased by a process which included the immersion of the blanks in ethyl alcohol in a Petri dish whose bottom was lined with open cell urethane foam, then scrubbing both sides of each blank with a foam swab. The crystals were then placed into the slots in the etching jig and agitated ultrasonically in a detergent solution, then rinsed thoroughly. Plasma cleaning in an oxygen plasma¹⁵ has been found to be a reliable alternative to scrubbing.

3. From the final rinse the crystals are transferred, while wet, into the etching bath, and are jigged around to make sure that there are no trapped air bubbles in the etching jig. During etching, the crystals are agitated in both directions to assure even etching on both sides. The approximate etching time can be calculated from the etching rate vs. temperature curve, Fig. 7, which will be discussed later.

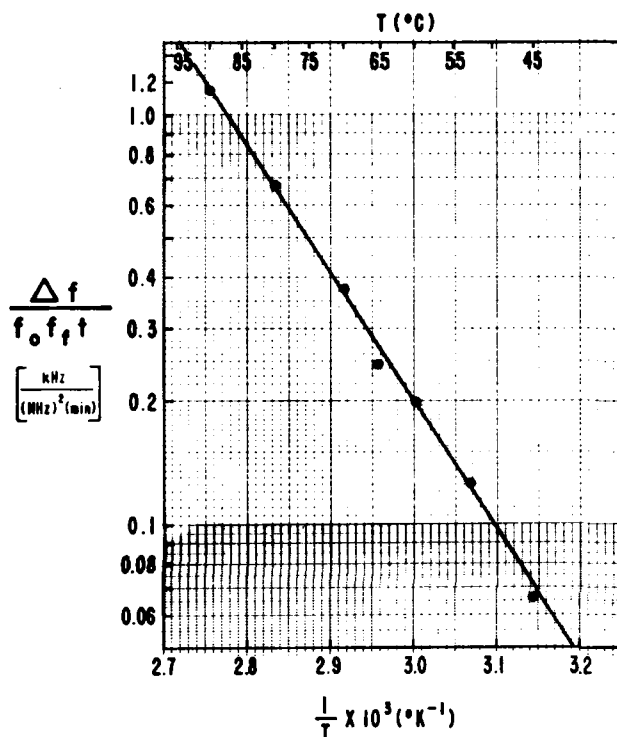


Fig. 7. Temperature dependence of etching rate.

4. After the crystals reach the desired frequency, the etching jig is removed rapidly from the etch bath and is immersed immediately into a container of hot water, given a thorough rinse under running hot water, then agitated ultrasonically in hot water, then given another rinse in running distilled water, then dried by spin drying. A thorough rinse is important in order to remove all residues of the etchant.

5. The crystals are then inspected under a microscope for uniformity of etch, and for defects such as scratch marks, etch pits and etch channels.

Non-uniform appearance of the etched surfaces can have the following causes: a. contamination on the crystals prior to etching, b. contamination in the etch bath which adsorbs onto the crystals, c. air bubbles on the crystals which remain after the crystals are immersed in the etch bath, d. contact between the crystals and the etching jig due to poor jig design or inadequate agitation, e. poor quality quartz in which a large number of etch pits and etch channels are produced by the etching, and f. inadequate rinsing at the completion of etching.

Silicon dioxide surfaces on oxidized silicon slices have been shown to retain fluorides even after

a two minute rinse in 95°C water¹⁶. The authors of this paper ascribe the "tenacious behavior of the fluoride ion...to the strength of the silicon-fluoride bond (136 kcal/mol compared to 106 kcal/mol for Si-O) on the ... SiO₂ surface". A careful analysis by Auger Electron Spectroscopy of several etched quartz blanks which had been rinsed thoroughly with distilled water failed to reveal any traces of fluorine on the surfaces. (To minimize the possibility of electron beam induced desorption of the fluorine, the samples were translated across the primary beam continuously while repeatedly scanning a narrow range of energies about the fluorine peak.)

Inspection Procedure

The inspection of etched blanks is performed under a microscope at about 40X magnification, with the light incidence perpendicular to the axis of the microscope, and with using a black background. First the crystal is inspected for surface irregularities such as scratch marks, pits and twinned areas by tilting the crystal so as to reflect light into the microscope. The crystal is then inspected for etch channels by holding it so that the light incidence is in the plane of the crystal (i.e., edge illumination). The etch channels are most visible when the edge illumination is incident along a direction perpendicular to the direction of the channels. For example, in most cultured quartz, the etch channels tend to be along directions near the Z direction. These channels are most easily visible therefore with the light incident from the X direction. To help make the etch channels more visible without rotating the crystals, it is helpful to use for the edge illumination two lights incident at a right angle to each other, or a ring light. The etch channels appear as small, bright streaks which extend through the blank from one side to the other. The thicker the blank, the longer the streaks, and the deeper the blank has been etched, the brighter the streaks.

Etching Variables

The etching variables investigated in this study were: the depth of etch, the surface finish prior to etching, the etching temperature, agitation during etching, etching bath depletion, and the quality of quartz used.

a. Depth of Etch

Figure 8 shows an SEM micrograph of the 3 μm alumina lapped surface prior to etching. The surface is filled with pits, crevices, and loose and nearly loose particles. After etching to Δf = 1.5f_of_f, the loose and nearly loose particles have been etched away, as shown in Fig. 9, however, numerous pits remain visible. A surface etched to Δf = 2.1f_of_f, Fig. 10, shows fewer pits. The pits have become broader and shallower.

Fig. 11 shows the topography corresponding to a depth of etch of Δf = 3.6f_of_f. Evidence of the surface damage produced by lapping has nearly disappeared. The hills and valleys, which tend to extend along about the X direction, have become more apparent. The surface roughness at this point is 0.19 μm, and the roughness angle is 2.9°.

As the etching progresses, the hills move further apart, and the valleys become shallower. Fig. 12 and 13 show the topographies corresponding to depths of etch of Δf = 8.9f_of_f and Δf = 31.5f_of_f. The surface roughnesses and roughness angles are 0.13 μm and 2.2°, and 0.09 μm and 0.7°, respectively.

Upon further etching, the topography changes less and less with increasing depth of etch. The surface

roughnesses and roughness angles approach the equilibrium values of $0.075 \mu\text{m}$ and 0.25° , respectively. Beyond a depth of etch of $\Delta f = 80f_{of}$, the topography no longer changes with increasing depth of etch. Fig. 14 shows an SEM micrograph for a surface etched to $\Delta f = 85f_{of}$.

The effect of the hills and valleys on the surface areas of these chemically polished blanks is small. For example, at a depth of etch of $\Delta f = 25f_{of}$, the ratio of surface areas between chemically polished blanks and "perfectly" polished blanks is less than 0.01%.



Fig. 8. Unetched $3 \mu\text{m}$ lapped surface.



Fig. 9. Surface etched to $\Delta f = 1.5f_{of}$.

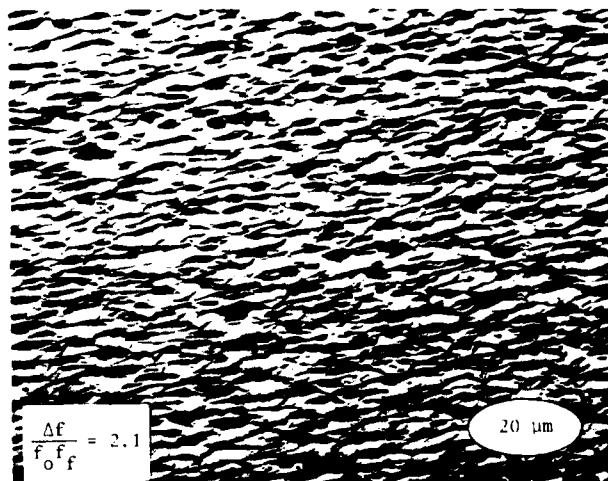


Fig. 10. Surface etched to $\Delta f = 2.1f_{of}$.

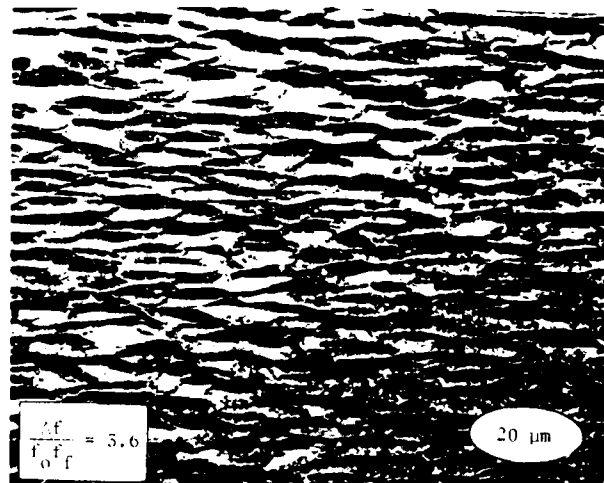


Fig. 11. Surface etched to $\Delta f = 3.6f_{of}$.

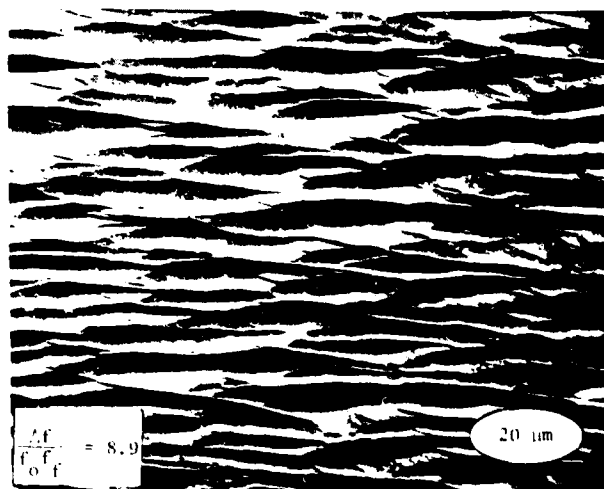


Fig. 12. Surface etched to $\Delta f = 8.9f_{of}$.

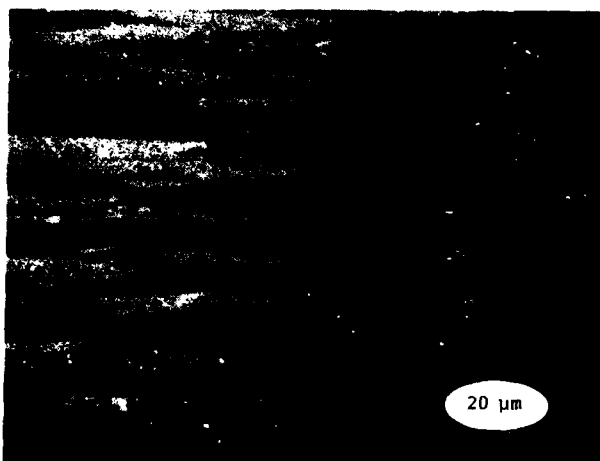


Fig. 13. Surface etched to $\Delta f = 31.5f_0f_f$.



Fig. 14. Surface etched to $\Delta f = 85f_0f_f$.

b. Surface Finish Prior to Etching

Four groups of crystal blanks, each with a different surface finish prior to etching, were etched at 85°C . The surface roughness and roughness angle vs. depth of etch was determined for each group. The four groups were prepared as follows: one group was cut with a diamond saw blade (80 grit) and was not lapped at all, the second group was final lapped with $12\text{ }\mu\text{m}$ MICROGRIT aluminum oxide abrasive, the third group was final lapped with $3\text{ }\mu\text{m}$ MICROGRIT, and the fourth group was lapped with $3\text{ }\mu\text{m}$ MICROGRIT, etched, then polished with cerium oxide.

The surface roughnesses and roughness angles of the blanks with the diamond saw blade surface finish continued to decrease, with little sign of leveling off, all the way to a depth of etch of $\Delta f = 300f_0f_f$, the maximum depth to which this group was etched. The surface roughnesses and roughness angles decreased nearly linearly from $3.0\text{ }\mu\text{m}$ and 3.6° , respectively, at $\Delta f = 110f_0f_f$, to $1.2\text{ }\mu\text{m}$ and 0.9° at $\Delta f = 300f_0f_f$. As the etching progressed, the blanks became more and more transparent.

The surface roughnesses and roughness angles of both the $12\text{ }\mu\text{m}$ and $3\text{ }\mu\text{m}$ lapped groups leveled off after a depth of etch of approximately $\Delta f = 80f_0f_f$, as shown

in Figs. 15 and 16. The equilibrium surface roughnesses are $0.45\text{ }\mu\text{m}$ and $0.075\text{ }\mu\text{m}$; the equilibrium roughness angles are 0.75° and 0.25° , respectively. The ratios of the equilibrium values for these two groups are of the same order as the ratios of the average particle sizes of the two abrasives used in the final lapping.

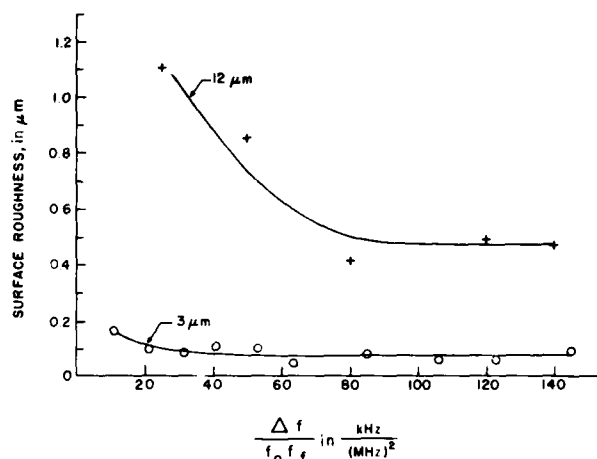


Fig. 15. Surface roughness vs. depth of etching for $3\text{ }\mu\text{m}$ and $12\text{ }\mu\text{m}$ lapped blanks.

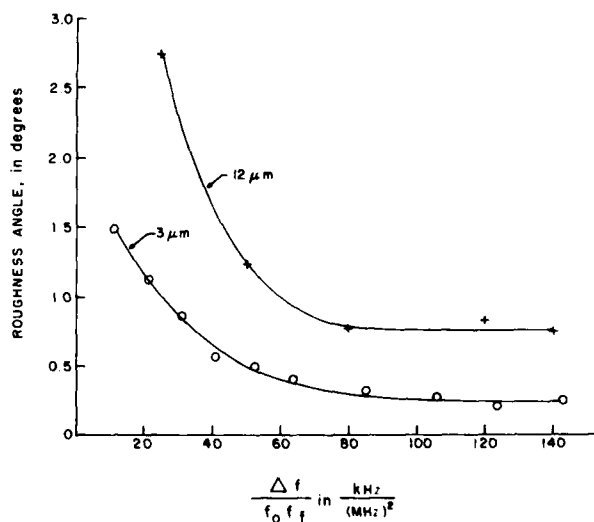


Fig. 16. Roughness angle vs. depth of etching for $3\text{ }\mu\text{m}$ and $12\text{ }\mu\text{m}$ lapped blanks.

The regular array of hills and valleys which developed on the lapped blanks did not develop on the group of polished blanks. The equilibrium topographies on these blanks consisted of smooth areas and scratch marks only. Fig. 17 shows the profile meter scan for a polished blank after it was etched to $\Delta f = 92f_0f_f$. It shows a $0.28\text{ }\mu\text{m}$ deep scratch mark, plus some smaller ones. However, between the scratch marks, the surface is smooth, to within the resolution of the profile meter, which is $0.01\text{ }\mu\text{m}$.

These results are consistent with the contention

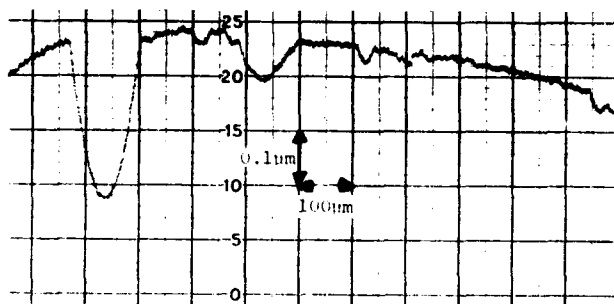


Fig. 17. Profile meter scan of polished blank after etching to $\Delta f = 92f_{of}$.

that the etching is diffusion controlled. The rougher the initial surface, the rougher the final equilibrium surface, where the depleted surface layer of etching solution has uniform thickness everywhere, and the hills therefore, are no longer etched faster than the valleys. A smooth, undisturbed initial surface remains smooth even after a large amount of material is removed by the etching. No signs of preferential etching along the different crystallographic axes appeared on the polished blanks.

The scratch marks represent lattice disturbances which had been produced by the mechanical polishing operation. The reason these regions are etched faster than the surrounding areas is the same as the reason for the formation of etch pits and etch channels, which will be discussed later in this paper.

The results above for the 12 μm and 3 μm lapped surfaces are not necessarily representative of surfaces lapped with the same size designation abrasives from different manufacturers. The typical size distribution curves for the 12 μm and 3 μm MICROGRIT WCA abrasives show the median particle diameters to be 11.75 μm and 2.85 μm respectively, and the maximum particle diameters to be approximately 30 μm and 10 μm , respectively¹⁷. The particles tend to have flat, platelet-like shapes. Fig. 18 shows a SEM micrograph of some 3 μm suspension treated MICROGRIT WCA particles. Abrasives from other manufacturers may have different particle size distributions and/or shapes, and may therefore produce different equilibrium surface topographies upon chemical polishing.



Fig. 18. 3 μm MICROGRIT abrasive.

The above results imply that the less deeply and the more uniformly disturbed the surface is prior to etching, the smoother will be the chemically polished surface. Accordingly, one should lap with as fine an abrasive prior to etching as possible. Since the finer the abrasive, the slower the lapping, it may be preferable to perform the chemical polishing in two steps in order to attain the optimum surface finish. First, to lap the crystals with a relatively coarse abrasive, e.g., 3 μm , then etch to $\Delta f = 10f_{of}$ to remove the damaged layer and chemically polish the surface. At this point the (r.m.s.) surface roughness is 0.11 μm , and the profile meter scans show that the largest hills on the surface (if the surface is free of etch channels) extend less than 0.5 μm above the valleys. To produce a semi-polished surface, it is therefore sufficient to remove $\Delta f \approx 0.3f^2$ with a fine abrasive. From such a semi-polished surface, a small amount of etching should produce a very smooth chemically polished surface.

In our experience, etching surfaces which had been fully polished with cerium oxide prior to etching have generally produced scratch marks. Although the above suggestion for using a semi-polished surface for chemical polishing has not yet been proven experimentally, the semi-polished surface may be preferable to a fully polished one, because rather than aiming for a perfectly scratch free surface, lapping with a fine abrasive produces a uniformly "scratched" surface similar to the ones produced by the 12 μm and 3 μm abrasives, except on a smaller scale.

c. Temperature

The temperature at which the etching is performed does not have a significant effect on the final surface roughness in the temperature range studied, 45°C to 90°C. The temperature does, however, have a significant effect on the rate of etching. Fig. 7 shows a semi-log plot of the temperature dependence of etch rate in saturated ammonium bifluoride. (Note the temperature scale in °C at the top of the graph.) The rate increases from $\Delta f = 0.063f_{of}$ per minute at 45°C, to $\Delta f = 1.1f_{of}$ per minute at 90°C. It is $1f_{of}$ per minute at 88°C, which implies that from a 3 μm surface, a chemically polished surface can be produced at this temperature in less than 15 minutes. The etch rate curve in Fig. 7 fits the equation

$$R(T) = Ae^{-\frac{E}{kT}}$$

where $R(T)$ is the etch rate in units of kHz per (MHz)·per minute, A equals $5.083 \times 10^8 \frac{\Delta f}{f_{of} \text{ min}}$, E (an

activation energy) equals 14.4 kcal/mol ($=0.62\text{eV}$), k is Boltzmann's constant, and T is the temperature in °K. All etch rates are the rates for chemically polished surfaces. (The initial etch rates on rough surfaces are always higher.)

The 14.4 kcal/mol activation energy is higher than the activation energies characteristic of diffusion controlled processes⁶, which are 4 to 6 kcal/mol. The simple expression above, however, does not take into account two effects. First, the etching solution is saturated at all temperatures, which means that the concentration increases with increasing temperature. Secondly, the etching reaction liberates heat which makes the temperature at which the reaction takes place higher than the bulk solution temperature. Both of these effects tend to change the apparent activation energy to a value which is higher than the "true" activation energy of the etching.

d. Agitation

The agitation rate has only a minor effect on the etch rate. For example, the etch rate with an agitation rate of 50 cycles per minute was compared with the etch rate with no agitation, at 85°C. A group of crystals was etched in 15 minute increments alternating between agitation and no agitation. The etch rates were found to be 11% higher during the agitation periods.

Such a change in etch rate at 85°C could be accounted for by a change in etchant temperature of only 2°C, which is close to the uncertainty in our temperature control. This uncertainty in temperature control could therefore possibly account for at least part of the observed increase in etch rate.

While agitation does not appear to have a significant effect on the etch rate, proper agitation, preferably in both directions, is important to assure that the crystals are etched uniformly on both sides. Agitation also serves to minimize temperature gradients in the etch bath, which in turn minimizes the blank to blank etch rate variations.

e. Etch Bath Depletion

To obtain an indication of the effects of a severe depletion of the etch bath, a handful (approximately 40 gms) of crystal blanks were dissolved completely in approximately 200 ml of etching solution. This is equivalent to the depletion that would be produced in a one liter etch bath by etching 100,000 blanks of 6.4 mm diameter to $\Delta f = 14f_{ff}$. The "depleted" etch bath contained enough $\text{NH}_4\text{F} \cdot \text{HF}$ flakes to assure that some undissolved flakes remained on the bottom of the bath after all the crystal blanks had dissolved.

A group of blanks was then etched in this bath from 14.8 MHz to 20.0 MHz ($\frac{\Delta f}{f_{ff}} = 17.6$) at 53°C. SEM

photos of the etched surfaces revealed no difference between these surfaces and surfaces etched the same amount in a fresh etching bath. The etching rate in the depleted etch bath, however, was lower by a factor of seven.

No attempt was made to chemically analyze the depleted etch bath. Judge¹⁸ has, however, examined by Raman spectroscopy a solution of NH_4F and HF ("7:1 buffered HF") before and after dissolving SiO_2 in it. The "before" solution showed no detectable absorptions. In the spectrum of the solution with SiO_2 dissolved in it, "a very large number of peaks are observed, indicating the presence of multiple product species". The lower etch rate in the "depleted" etch bath is probably due to the presence of a high concentration of these species.

f. Quality of Quartz - Natural vs. Cultured

At the completion of chemical polishing, the blanks are inspected for etch pits and etch channels using the technique described earlier. At the two points where an etch channel intersects the blank surfaces etch pits have always been found, however, the etch pits are not always associated with etch channels. Although the effects of etch pits and etch channels on resonator performance have not yet been determined, we are currently rejecting all blanks with etch channels. There are primarily two reasons for this. First, the etch channels are probably filled with ammonium bifluoride at the completion of etching. Rinsing the blanks will almost certainly fail to remove all of the fluorides from the channels. The presence of fluorides

in the etch channels, therefore, is likely to adversely affect the resonator's stability. Since the etch rate through the etch channels can be about 100 times as high as through the rest of the blank, even lightly etched blanks may contain channels which extend at least part way into the blanks.

The second reason is that the inspection for etch channels can be performed early in the resonator fabrication sequence. For example, in the fabrication of circular resonators, a deep etch and inspection for channels can follow the rounding operation, at which point the cost of a blank is still relatively small. The deep etch at that point can also serve to prepare the blanks for fine lapping by removing material from the blanks rapidly, conveniently, without angle change, and without producing any surface damage. (In fact, the deep etch removes the surface damage left by the previous abrasive.)

Since the institution of the deep etch followed by inspection as part of our resonator fabrication process, we have inspected well over one thousand natural quartz blanks, in several lots. The blanks were 6.4 mm diameter, and were intended for being made into 20 MHz fundamental resonators. The incidence of etch channels in these 6.4 mm diameter blanks was between 30% and 40%, i.e., in each lot, 60% to 70% of the blanks were free of etch channels. In larger diameter blanks, of course, the incidence of etch channels is higher.

When the chemical polishing was attempted on a group of cultured quartz blanks, the resulting surfaces appeared to the unaided eye to be unevenly etched. A closer examination of the surfaces revealed that the poor chemical polish was due to a very large number of etch pits on the surfaces, most of which were the terminations of etch channels. Figs. 19, 20 and 21 show SEM micrographs at different magnifications of one such surface showing the etch pits and etch channels. None of the blanks in this group were free of etch channels.



Fig. 19. A cultured quartz blank after etching.



Fig. 20. Etch pits in cultured quartz.



Fig. 21. Etch channel at the bottom of an etch pit.

To investigate how the different varieties of cultured quartz compare with natural quartz with respect to the incidence of etch channels, sixteen varieties of cultured quartz from ten different growers were deeply etched and inspected for etch channels. Three of the varieties were swept quartz.

The sixteen varieties could be divided into three categories. One variety, a swept quartz, was free of etch channels. Five varieties had only a few etch channels per blank - on the order of 10 channels per cm^2 of blank surface. The two remaining swept varieties

were in this second category. Fig. 22 shows a light micrograph of how such a blank appears during inspection with edge illumination. The remaining ten varieties contained a large number of etch channels per cm^2 . Inspection of blanks from the worst varieties in this group is comparable to observing the Milky Way at night - as shown in Fig. 23.

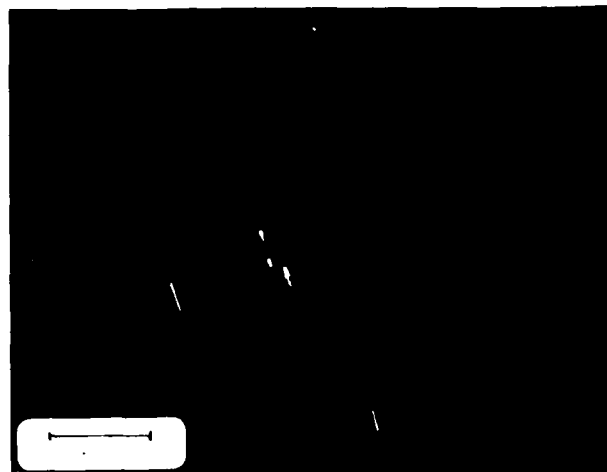


Fig. 22. Etch channels observed under a microscope with edge illumination.

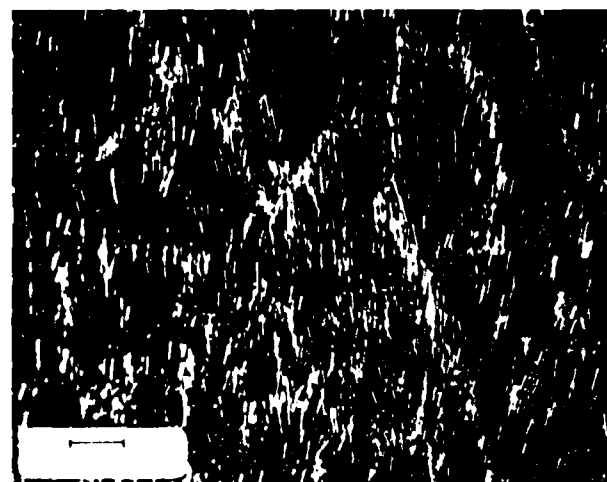


Fig. 23. Etch channels in a blank made of poor quality cultured quartz.

The one variety which was free of etch channels was vacuum swept at Sandia Laboratories by a process similar to that described by King¹¹. Some of the effects of sweeping in vacuum, as opposed to sweeping in other atmospheres, has also been discussed by Krefft¹², and by King and Sander²¹, who showed that resonators made from vacuum swept cultured quartz exhibited much lower frequency offsets after burst irradiation than resonators made from natural quartz or air swept cultured quartz.

We were able to obtain a few blanks from both the

vacuum swept and unswept portions of the same bar of cultured quartz. Whereas the blanks from the swept portion were completely free of etch channels (three blanks were etched, each 15 mm in diameter) the blanks from the unswept portion contained numerous etch channels. Fig. 24 shows a side by side comparison of the swept and unswept blanks as they appeared under the microscope with edge illumination. Even after the swept blanks were etched further, until a total of 300 μm was removed in an 85°C etch bath, not a single etch channel was observable. There were, however, several etch pits visible on each of the surfaces. The number of etch pits on the vacuum swept blanks appeared to be fewer than on the unswept blanks, although the sample sizes were not large enough to allow us to draw a definite conclusion about this, especially since there were areas on both types of blanks which were free of etch pits.



Fig. 24. Vacuum swept vs. unswept blanks from the same bar of cultured quartz.

Etch channels in quartz have been observed previously²²⁻²⁹. They are probably a consequence of dislocations at which impurities had segregated²². Impurities dissolved in the quartz, particularly those which do not fit into the lattice, segregate at dislocations. There is a strain energy associated with the presence of these impurities which results in the dislocation being etched more readily than the rest of the surface, i.e., the activation energy for etching is reduced by the strain energy. The dislocation may also act as a point of nucleation of attack. In such a case, when there is also a considerable heat of reaction, the heat generated can enhance both the diffusion and the rate of reaction in the vicinity of the dislocation, which can result in an etch channel and/or etch pit being formed. Fig. 25 shows an SEM micrograph of etch pits on the surface of a deeply etched ($\Delta f \approx 85f_0 f_f$) natural quartz blank. In natural quartz, some of the etch channels may be due to fission tracks¹⁰.

Sweeping is known to be able to remove interstitial impurities from quartz. The low incidence of etch channels in the three swept quartz varieties is probably due to the relief of the strain energy associated with these impurities. The difference in the incidence of etch channels between the vacuum swept quartz and the other two swept varieties (which were swept in air) may possibly be accounted for by the fact that when sweeping takes place in air, as the impurities in the crystal diffuse toward the cathode, impurities from the air can diffuse into the crystal from the anode side²⁰.

The presence of etch channels also seems to be associated with an apparent increase in the etch rate of

quartz. A group of natural quartz and a group of one of the better quality cultured quartz blanks were etched and inspected for channels. From the group of natural quartz blanks the ones with etch channels were eliminated. The cultured quartz blanks were 7.6 mm diameter and contained about ten channels per blank. The two groups of blanks were etched simultaneously side by side, in the same etch bath, at 85°C. The rate of etching was measured, as usual, by measuring $\frac{\Delta f}{f_0 f_f}$ per minute.

It was found that the etch rate for the cultured quartz group was 1.6% higher. When the experiment was repeated with a group of natural quartz blanks without any etch channels and a group of poor quality, "Milky Way-like" quartz, the etch rate of the "Milky Way" quartz was found to be 4.9% higher. When two groups of natural quartz blanks were etched side by side, although the groups had been obtained from different suppliers and had different initial surface finishes (12 μm vs. 3 μm), different frequencies (6.5 MHz vs. 5 MHz), and different geometries (circular vs. square), the two etch rates differed by less than 0.2%. (All etch rates are measured on chemically polished blanks.)



Fig. 25. Etch pits in a deeply etched natural quartz blank.

The Strengths of Chemically Polished Blanks

The theoretical inherent strengths of elastic materials are generally orders of magnitude greater than the measured strengths of the ordinary forms of these materials. This reduction in strength is known to be caused principally by surface flaws, such as scratches, which concentrate the applied stress and thereby lead to fracture at loads which are much lower than the theoretical maximum. Removal of all flaws from the surface permit the strengths of materials to approach the theoretical maximum. For example³¹, the theoretical strength of fused quartz (and of other glasses) is about $1.8 \times 10^{10} \text{ N/m}^2$. The actual strengths measured are in the range of 10^7 to 10^8 N/m^2 . However, by heating glasses to above their softening point (i.e., fire polishing) the surface flaws can be eliminated. The strength of fire polished fused quartz has been measured to be as high as $1.5 \times 10^{10} \text{ N/m}^2$, nearly as high as the theoretical maximum.

Since the surfaces of chemically polished quartz blanks appear to be free of surface damage, an experiment was performed to determine how the strengths of these blanks compare with the strengths of mechanically polished blanks and with the strengths of lapped blanks. The experiment consisted of supporting the blanks only at the edge, placing a spherical

indenter at the center of the blank, gradually increasing the load on the indenter, and measuring the load required to fracture the blank. A photograph of the apparatus, together with a sketch of the area where the indenter contacts the blank, is shown in Fig. 26. The indenter was a steel ball which was connected by means of a shaft to a platform on which a 2000 ml plastic beaker could be placed. The load on the indenter was increased by pouring water into the beaker. The blank was supported by a 0.5 mm wide rim made of a hard plastic.

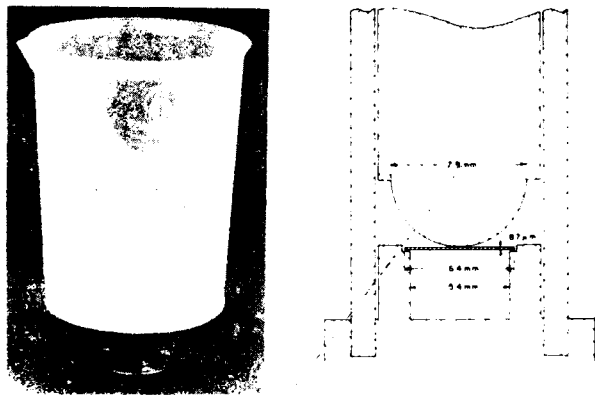


Fig. 26. Apparatus for blank strength evaluation.

A group of twelve 20 MHz, 6.4 mm diameter AT-cut blanks were tested. The blanks had been etched $15f_{off}$, from a 3 μ m lapped initial surface. The average load on the indenter at breakage was 2.2 kg. Two of the blanks did not break when the beaker was overflowing with water - a load of 3.1 kg. (These numbers should be considered to be only a semi-quantitative indication of the inherent strength of these blanks, because for example, each time a blank fractured, the rim on which the blanks were supported was damaged. The load on the next blank, therefore, may not have been as uniformly distributed.)

It was not possible with this apparatus to measure the strengths of blanks of the same dimensions which had been mechanically polished or lapped with 3 μ m abrasive, because all such blanks broke due to the weight of the platform alone (i.e., without the beaker) - a weight of 157 gms.

To place the strength of the chemically polished blanks in perspective, consider the fact that the weight of a 20 MHz blank of 6.4 mm diameter is only 7.0 mg; i.e., the ratio between the average load on the indenter at breakage and weight of the blank is about 300,000. This also means that the shock level at which the load on the blank due to its own weight is equal to the average load on the indenter at breakage is 300,000g.

To obtain a rough approximation of the shock levels the 20 MHz fundamental chemically polished blanks can survive when supported by four point mounts, the rim on which the blanks were supported was reduced to four small areas 90° apart. The dimensions of each area were 0.5 mm in the radial direction and 0.4 mm in the tangential direction. The average load on the indenter at breakage was thus reduced to 550 gms, which corresponds to a load due to a shock of about 80,000g.

An additional indication of the high resistance to fracture in these chemically polished blanks is that when some 14 mm diameter blanks were etched from 15 MHz

to above 100 MHz, the blanks flexed like very thin sheets of mica. For example, holding such a blank at diametrically opposite points between two fingers the blank could be bent about 90° before it broke.

The Q of Chemically Polished Resonators

The Q of a resonator can be defined by

$$Q = 2\pi \frac{\text{Energy Stored}}{\text{Energy Dissipated per Cycle}}$$

The energy dissipated per cycle is the sum of the energies dissipated by the various loss mechanisms. For a chemically polished resonator, the losses may be due to: the mounting, the bonding, the electrodes, contamination, the atmosphere surrounding the resonator, the intrinsic losses in the quartz, plus the losses due to the features produced by chemical polishing, i.e., the regular array of hills and valleys, the etch channels and the etch pits. Q^{-1} can therefore be expressed as the sum of the Q^{-1} 's due to the various loss mechanisms, i.e.,

$$\frac{1}{Q} = \frac{1}{Q_{\text{mounting}}} + \frac{1}{Q_{\text{bonding}}} + \dots + \frac{1}{Q_{\text{etch pits}}}$$

If the regular array of hills and valleys due to chemical polishing produces an inherent Q degradation, then in a group of resonators made from blanks which received the same chemical polishing, the resonator with the highest Q represents a lower limit on the Q achievable with that chemical polish (at the given frequency).

Since these investigations on chemical polishing began, we have fabricated approximately 300 resonators using chemically polished blanks. All of the resonators were fundamental mode, in the 18 MHz to 22 MHz range. The blanks were plano-plano, 6.4 mm diameter natural quartz. Most of the resonators were of the high shock, nickel electrobonded type described previously³². The depths of etch ranged from $\Delta f = 2f_{off}$ to $\Delta f = 22f_{off}$. The surfaces prior to etching had been lapped with 3 μ m MICROGRIT. For the highest Q resonators in each group, the motional capacitances ranged from 12 fF to 13.5 fF, the resistances ranged from 3 Ω to 5 Ω , and the Q's thus ranged from 140,000 to 210,000, with no apparent Q degradation with depth of etch.

The highest Q resonators made from polished blanks of the same blank geometry, but of a low shock nickel electrobonded design³³, had Q's which were no higher ($Q = 180,000$) than the highest Q resonators made from chemically polished blanks. In fact, the highest Q resonator of that blank geometry we have made to date ($Q = 210,000$) was a 19 MHz high shock resonator, with a depth of etch of $\Delta f = 16f_{off}$.

In an attempt to investigate the effect of an extreme amount of etching, a group of crystal blanks were etched from 10.5 MHz to 22.4 MHz, which corresponds to a $\Delta f = 50f_{off}$. Prior to etching, these blanks had been in process to be made into 20 MHz resonators. Normally, the blanks would have been lapped to at least 15 MHz with 3 μ m MICROGRIT before being etched to the final frequency. Although the 10.5 MHz blanks were lapped with the 3 μ m abrasive, the amount of material removed with this abrasive was much less than usual, and the surface finish prior to etching may not have been a true 3 μ m finish.

The highest Q resonator in this group (of eleven

resonators) had a Q of 110,000 ($C_1 = 13.4$ fF, $R = 5\Omega$). Most of the resonators in this group had anomalously low C_1 's, the cause of which has not yet been determined. This experiment did show, however, that if the extremely deep etching produces a Q degradation, this degradation is not a drastic one - at least not for fundamental mode resonators in the 20 MHz range. The Q's of the above resonators were not evaluated on overtone operation because the electrodes were gold, with about a $1.0f^2$ plateback, which is known to be too heavy for overtone operation³⁴.

As the frequency increases so that the wavelength of the acoustic wave approaches the dimensions of the etch features, it would not be surprising to find a Q degradation. However, in any systematic study of the inherent Q degradation due to the array of hills and valleys which develop during chemical polishing, great care would have to be taken to use only those blanks which are free of both etch channels and etch pits, since the dimension of these features are usually at least an order of magnitude larger than the dimensions of the hills and valleys. In the above resonators, the blanks selected were free of etch channels, but not necessarily of etch pits.

Conclusions

Etching in a saturated solution of ammonium bifluoride has been shown to be capable of chemically polishing AT-cut quartz over a broad range of conditions. The quality of surface finish which can be produced by this method depends primarily on the surface finish prior to etching, the depth of etch and the quality of quartz used. The speed at which chemically polished surfaces can be produced depends primarily on the temperature of the etch bath. In an 88°C etch bath, for example, starting with 3 μ m lapped blanks, chemically polished surfaces can be produced in less than fifteen minutes. Depletion of the etch bath slows down the etching rate but does not affect the surface roughness.

The process can remove large amounts of material from lapped blanks while simultaneously producing an improved surface finish, without producing shifts in the angles of cut. The process can also produce blanks of great strength, which is particularly important for high shock applications.

Since depth of etch is not critical, the chemical polishing may allow manufacturers to stock lapped blanks at only a few frequencies at each commonly used angle of cut, then etch the blanks to the required frequencies as the need arises. The method may also permit the manufacturing of miniature contoured high frequency resonators, since such small diameter resonators can now be contoured at conventional frequencies, then etched up to high frequencies. If a masking material which is resistant to the etch solution can be found, the method may also permit the fabrication of high frequency resonators and filters with the inverted mesa structure.

Inspection of deeply etched blanks for etch channels is a simple screening technique which can reveal large differences among blanks made from different varieties of quartz. The incidence of etch channels has not yet been correlated with resonator performance parameters such as stability, Q, radiation hardness, etc. From indirect evidence, however, it seems probable that, everything else being equal, resonators made from blanks which are free of etch channels will perform better than those which show etch channels upon deep etching.

Acknowledgements

The authors gratefully acknowledge the skillful contributions of: D. Eckart, C. F. Cook and A. Dunlap who provided the SEM and light micrographs, J. Falcone who provided the profile meter scans, D. Fox who performed the Auger Electron Spectroscopy, and H. Spaight who prepared some of the lapped blanks used in these experiments. We would also like to thank our friends in the crystal industry who provided us with the samples of different varieties of cultured quartz. We are particularly grateful to Tom Young of Sandia Laboratories for supplying the vacuum swept and unswept blanks from the same bar of cultured quartz.

References

1. Rabinowitz, E., "Polishing", Scientific American, Vol. 218, pp.91-99, June 1968.
2. Holland, L., "The Properties of Glass Surfaces", London: Chapman and Hall, 1964, Chapter 1.
3. Newkirk, J.B., Young, J.D. and Spencer, J.P., "Defect Structures on Polished Quartz Surfaces as Seen by X-ray Diffraction Microscopy", Jour. Appl. Phys. Vol. 35, pp.1362-63, (1964).
4. Frondel, C., "The System of Mineralogy . . .", Vol. III, Silica Minerals, New York and London: John Wiley and Sons, Inc., pp.162-163, (1962).
5. Ernsberger, F.M., "Structural Effects in the Chemical Reactivity of Silica and Silicates", Jour. Phys. Chem. Solids, Vol. 13, pp. 347-351, (1960).
6. Holmes, P.J., "The Electrochemistry of Semiconductors", London and New York: Academic Press, Chapters 6 and 8, (1962).
7. Micro Abrasives Corp., 720 Southampton Rd., Westfield, MA 01085.
8. Talysurf 4, The Rank Organisation, Rank Taylor Hobson, Leicester, England.
9. Model 9869A Digitizer with a 9820A Calculator, Hewlett-Packard Co., Palo Alto, CA 94303.
10. Model No. J40, Fluoroware, Inc., Jonathan Industrial Center, Chaska, MN 55318.
11. Model No. 4812, Cole Parmer Co., 7425 N. Oak Park Ave., Chicago, IL 60648.
12. Model No. A14-01. Fluoroware, Inc.
13. Catalog No. A-664, Fisher Scientific Co., Fair Lawn, NJ 07410 - also obtained the same results with Catalog No. 0618, J. T. Baker Chemical Co., Phillipsburg, NJ 08865
14. Guide to Physical Properties of J. T. Baker Industrial Inorganic Chemicals, J.T. Baker Chemical Co., Phillipsburg, NJ 08865.
15. Plasmod model, Tegal Corp., 860 Wharf St., Richmond, CA 94804.
16. Larrabee, G. B., Heinen, K. G. and Harrel, S.A., "Measurement of the Retention of Fluoride by Silicon and Silicon Dioxide Surfaces", Jour. Electrochem. Soc., Vol. 114, pp. 867-869, August 1967.
17. The particle size analysis was performed on a Coulter Counter, Model T, Coulter Electronics Inc., 590 W.

- 20th St., Hialeah, FL 33010 - data supplied by Micro Abrasives Corp. - see ref. no. 7.
18. Judge, J.S., "A Study of the Dissolution of SiO_2 in Acidic Fluoride Solutions", *Jour. Electrochem. Soc.*, Vol. 118, pp. 1772-75, November 1971.
 19. King, J.C., "Vacuum Electrolysis of Quartz", US Patent No. 3,932,777, Jan. 13, 1976.
 20. Krefft, G. B., "Effects of High Temperature Electrolysis on the Coloration Characteristics and OH-Absorption Bands in Alpha-Quartz", *Radiation Effects*, Vol. 26, pp. 249-259, (1975).
 21. King, J.C. and Sander, H.H., "Transient Change in Q and Frequency of AT-cut Quartz Resonators Following Exposure to Pulse X-rays", *IEEE Trans. on Nuclear Science*, Vol. NS-20, pp. 117-125, (1973).
 22. Nielsen, J.W. and Foster, F.G., "Unusual Etch Pits in Quartz Crystals", *The American Mineralogist*, Vol. 45, pp. 299-310, March-April 1960.
 23. Arnold, G.W., Jr., "Defects in Quartz Crystals", *Proc. of the 11th Annual Symposium on Frequency Control*, US Army Electronics Command, Ft. Monmouth, NJ, pp. 112-129. Copies available from NTIS, AD 298323.
 24. Augustine, F., "Improving the Quality of Synthetic Quartz", *Proc. of the 11th Annual Symposium on Frequency Control*, US Army Electronics Command, Ft. Monmouth, NJ, pp. 130-141, (1957).
 25. Augustine, F. and Hale, D.R., "Topography and Etch Patterns of Synthetic Quartz", *Jour. Phys. Chem. Solids*, Vol. 13, pp. 344-346, (1960).
 26. Hanyu, T., "Dislocation Etch Tunnels in Quartz Crystals", *Jour. Phys. Soc. Japan*, Vol. 19, p. 1489, (1964).
 27. Gülzow-Qual, G., "Ätzversuche an Synthetischem Quarz", *Physica Status Solidi*, Vol. 1, pp. 62-67, (1961).
 28. Tsinzerling, E.V. and Mironova, Z.A., "Revealing Dislocations in Quartz by the Selective Etch Method", *Sov. Phys. Crystallogr.*, Vol. 8, pp. 89-91, (1963).
 29. Volskaya, O.B., "Selective Etchants for Quartz", *Sov. Phys. Crystallogr.*, Vol. 13, pp. 620-623, Jan-Feb 1969.
 30. Macdougall, J.D., "Fission - Track Dating", *Scientific American*, Vol. 235, pp. 114-122, Dec 1976.
 31. Doremus, R. H., "Glass Science", New York: John Wiley & Sons, Inc., Chapters 15 and 17, (1973).
 32. Filler, R. L. and Vig, J.R., "The Effect of Bonding on the Frequency vs Temperature Characteristics of AT-cut Resonators", *Proc. of the 30th Annual Symposium on Frequency Control*, US Army Electronics Command, Ft. Monmouth, NJ, pp. 264-268, (1976). Copies available from Electronics Industries Assoc., 2001 Eye Street, NW, Washington, DC 20006, also available as AD A030316.
 33. Vig, J.R., LeBus, J.W. and Filler, R.L., "Further Results on UV Cleaning and Ni Electrobonding", *Proc. of the 29th Annual Symposium on Frequency Control*, US Army Electronics Command, Ft. Monmouth, NJ, pp. 220-229, (1975). Copies available from EIA, as in Ref. 32.
 34. Bennett, R. E., editor, "Quartz Resonator Handbook", Union Thermoelectric Div., 1960, pp. 113-115. Copies available from NTIS, AD 251289.

AGING ANALYSIS OF QUARTZ CRYSTAL UNITS WITH Ti Pd Au ELECTRODES

G. L. Dybwad
Bell Telephone Laboratories, Incorporated
Allentown, Pennsylvania 18103

Abstract

The observed frequency shifts with long storage times at several isothermal temperatures indicate that diffusion of extraneous material into the Ti Pd Au thin film electrodes controls the aging process of 8 MHz AT-cut quartz crystal units. Crystal units were fabricated using electron beam deposited Ti Pd Au electrodes, sealed in vacuum, and placed in isothermal aging (80°C, 140°C, 160°C) for over 8000 hours. The frequency behavior versus time was found to be complex in that the frequency first rose and then fell as time progressed. A non-linear, least squares fitting computer program was used to analyze the aging data. Several mathematical aging models were used in the program to fit the experimental data. The best fit involved two separate mechanisms: the first was a fast (~2-10 hour time constant) saturating process, probably desorption; the second was a slower process proceeding as $(\text{time})^{1/2}$. The latter result suggests that the physical frequency aging mechanism was associated with diffusion of extraneous material into the Ti Pd Au electrodes. The activation energy for this process was found to be 0.8 ± 0.2 eV. This energy is consistent with other measurements of the activation energy for diffusion in Ti Pd Au thin films.⁸ With the identification of the aging mechanisms as contamination oriented, we now understand why changes in the process have improved performance: electrode evaporation proceeds under higher vacuum conditions, and extra cleaning is applied to all units before sealing. The purpose of the higher vacuum is to reduce the density of film defects, thus producing electrodes which are more impervious to residual contamination. The extra cleaning reduces the post seal contamination levels which reduces the aging frequency shifts.

In general, aging data, when analyzed using mathematical models having physical interpretations, offers the device and process designer specific information as to which fabrication processes can be changed in order to alter device performance.

Introduction

Early isothermal aging (80, 140 and 160°C) results showed that 8 MHz crystal units with Ti Pd Au electrodes had complex frequency shifts (Figure 1).¹ Because the average frequency first rose, but then fell to negative values, it was clear that the standard technique of plotting linear frequency shift versus the logarithm of time in order to determine end-of-life performance by extrapolation would give incorrect estimates. This paper reports the computer analysis of the complex Ti Pd Au aging. The mathematical analysis used formulae which had physical interpretations. The formula giving the best fit then gave us insight into the physical mechanisms causing the complex behavior. Performance improvements from prior assembly process changes then became clearly understood.

Experimental Procedure

Crystal Units

The AT-cut crystal plate blanks (5 μ m lapped finish, etched 100 kHz and cleaned) were loaded into

Sloan deposition fixtures.² The fixtures were equipped with shadow masks to define the electrode patterns. The fixtures were then placed in an electron beam evaporator (Temescal FC-1800) which was "dry pumped" using a high speed turbomolecular pump. Pressures were maintained between 10^{-6} and 4×10^{-5} torr during the deposition. An automatic rate controller (Kronos, Inc.) regulated the deposition. The thickness of each film layer was held constant to within 2% at:

Ti : 1,500 \AA
Pd : 500 \AA
Au : 3,500 \AA

After deposition was complete on one side, vacuum was broken, the fixtures were turned over, and the same procedure was carried out on the second side of the quartz blank. Completed units were unloaded, and leads were thermocompression bonded to the electrode tabs. Finally, each unit was vacuum sealed at about 10^{-5} torr in a metal can having feed-through leads.

Test Set

A previously described¹ computer controlled test set was used to make the frequency measurements. The reproducibility of measurements was approximately ± 0.5 Hz. All measurements were made at $24 \pm 2^\circ\text{C}$, and the initial measurement before starting the experiment was considered as zero time and zero frequency.

Aging Ovens

Commercial ovens (Fisher and A.T.L.) were used to elevate the temperature of the units. Temperature control was typically $\pm 2^\circ\text{C}$ for all temperatures (80, 140 and 160°C).

Computer Program

An algorithm developed by Marquart³ was used to curve fit the isothermal aging data. The program contains non-linear least squares curve fitting routines making it ideal for the complex curves exhibited in these experiments. The user forms a function with up to 30 unknown coefficients. The various partial derivatives of the function are also included in the main program. Starting with an initial "guess" set of values for the coefficients, the algorithm solves matrix equations, and calculates the sum of the squares of the residuals. The program then tries to minimize this sum to within a user defined, arbitrarily small number, ϵ (5×10^{-7} in these experiments). The algorithm performs another iteration until the error range ϵ is satisfied. The list of coefficients of the function giving the best fit is then printed out.

We tried several mathematical models which could be identified with known physical processes. These are listed in Figure 2. Equation 1 gives the standard log t plot; a straight line on linear-log paper. Equation 2 can be derived on the basis of chemical arguments of atoms on surfaces;⁴ as can be seen, two such mechanisms with different time constants are proposed here. One to satisfy the early positive Δf shifts and the second to describe the later negative shifts. Equation 3 assumes that the two mechanisms act like an

ordinary charging and discharging capacitor with time constants of A_2 and A_4 respectively. Equation 4 uses the capacitor model for the first mechanism and a power law for the second. Equation 5 is similar to Equation 4 but fixes the power law exponent at one-half. In other experiments by other investigators, it has been shown that such a form often describes diffusion of atoms in solids.

Results of Analysis

Figure 3 shows isothermal data for three temperatures. Each data point represents the average frequency shift of 9 units. The high temperatures accelerate the aging mechanisms greatly. The 160°C data is especially good for comparison of the curve fitting models because of the accentuated curvature. Figure 4 compares the fit obtained for the various models for the 160°C data. For this and the data at 80° and 140° the model which gave the best fit was model 4 in Figure 2; the charging capacitor followed by a power law. Figure 5 shows graphically the best fit curve for the 160°C data.

Capacitor Model

The time constant for this process (A_2) was a few hours at 25°C with an ultimate frequency shift (A_1) of about 0.5 ppm. More precise values for the capacitor model are lacking because only a few measurements were taken before saturation occurred and also, any room temperature shifts occurring after sealing and before the start of these experiments (~5 days) were not monitored. Such values would be consistent with a redistribution of residual contamination within the sealed enclosure. Since the resonator frequency shifted positively, contamination (~1 monolayer) tended to leave the electrode surfaces.

Power Law Model

This second mechanism dominates the overall crystal unit stability. Hence, understanding the controlling physical mechanism may suggest ways of controlling the unit stability.

Figure 6 lists the best fit exponent values, A_4 in Equation 4. The average value is very close to one-half. Theoretical calculations^{6,7} and experimental evidence⁸⁻¹⁰ have shown that diffusion of contaminants through a thin film is governed by a $t^{0.5}$ law. Since in our experiments frequency shifts were negative, we propose that residual, electrode surface contamination diffuses into the Ti Pd Au films while additional container contamination deposits on the Au surface. Figure 7 shows this model schematically. Auger electron spectroscopy confirms that surface oxygen levels correlate with the magnitude of observed frequency shifts.¹¹ However, the absolute surface mass of oxygen could not account for all the added mass needed to produce the observed negative frequency shifts. Rutherford backscattering experiments showed that aged Ti Pd Au crystal units show very little interdiffusion of the Ti and Pd metallic layers. Inhibition of interdiffusion typically occurs when metallic interfaces have become oxidized.¹² Hence lack of interdiffusion across the Ti/Pd and Pd/Au interfaces in our experiments supports the idea that surface contamination is gettering by the titanium and/or palladium layers.

Additional support for the idea of diffusion controlled frequency shifts for Ti Pd Au units comes from the observed activation energy in these experiments. An Arrhenius plot¹ of A_4 from equation 4 gave a value of 0.8 ± 0.2 eV. Diffusion of Ti Pd Au thin film conductors has been shown to be activated at about 0.9 eV

by Speight et. al.⁸ and Lau¹³.

Operating Procedures

We observed in our laboratory an improvement of Ti Pd Au aging performance when

1. extra care was taken to clean units prior to sealing, and
2. films were deposited under improved vacuum conditions in the evaporator.

The computer analysis of aging data followed by physical modelling of the responsible mechanisms has explained why these process changes have been effective. As the model shows, if there is no contamination sealed within the unit enclosure, nothing can diffuse into the films to cause large negative frequency shifts. Likewise, the improved deposition conditions give fewer film defects^{14,15} reducing the number of diffusion paths greatly. That is, under the improved conditions, the top gold layer has the desired protection properties.

Summary

We have described here another tool that the device designer can use (Fig. 8). Aging results when analyzed by the powerful and readily available Marquart curve fitting routine can yield physical insight into device instabilities. This understanding can be fed back to the responsible fabrication process step. The effect of changes in a process step can then be monitored by another iteration of the aging study.

Or, the analysis procedure can serve as a process control check. That is, having identified acceptable aging mechanisms, the curve fitting routine could identify new and unwanted mechanisms in product should they occur.

Specifically, we have identified a diffusion process in Ti Pd Au electrode films using this technique.

Acknowledgments

The author would like to thank R. C. Sundahl, Jr. for helpful discussions on models and suggesting the $t^{0.5}$ law, T. R. Meeker for aging data and R. C. Rennick for writing the user program for the Bell Telephone Laboratories computer system.

References

1. S. H. Olster et. al., 29th Annual Symposium on Frequency Control, May, 1975, p. 105.
2. H. F. Cawley, et. al., 29th Annual Symposium on Frequency Control, May, 1975, p. 113.
3. D. W. Marquardt, J. Soc. Indust. Appl. Math., **11**, 431 (1963).
4. W. L. Paschke, internal communication, 1964.
5. Bardeen and Herring, Imperfections in Nearly Perfect Crystals, Wiley and Sons, N.Y., N.Y., 1952.
6. J. Crank, The Mathematics of Diffusion, Clarendon Press, Oxford, England, 1975.
7. W. W. Mullins, J.A.P., **28**, No. 3, 333 (1957).
8. J. D. Speight et. al., IEEE 4th Proc. Reliability Physics, Las Vegas, Nevada, March 1971, p. 195.
9. J. M. Poate et. al., "Thin Film Interdiffusion I. Au-Pd, Pd-Au, Ti-Pd, Ti-Au, Ti-Pd-Au, and Ti-Au-Pd, J.A.P., **46**, 4275 (1975).
10. Seltzer and Jaffe, Defects and Transport in Oxides, Plenum Press, N.Y., N.Y., 1974, p. 390.
11. J. Bindell, Internal Communication, 2/27/75.

12. J. Beynon and L. Olumekor, *Thin Solid Films*, **41** (1977) L1-L2.
13. S. S. Lau and R. C. Sun, *Thin Solid Films*, **10**, 273 (1972).
14. Y. Namba and T. Mori, "Effect of Residual Gas on the Film Growth Process", *Shinko (Vacuum)*, **16**, No. 3, 96 (1973).
15. J. R. Lloyd and S. Nakahara, *J. Vac. Sci. Tech.*, **14**, No. 1, 655 (1977).

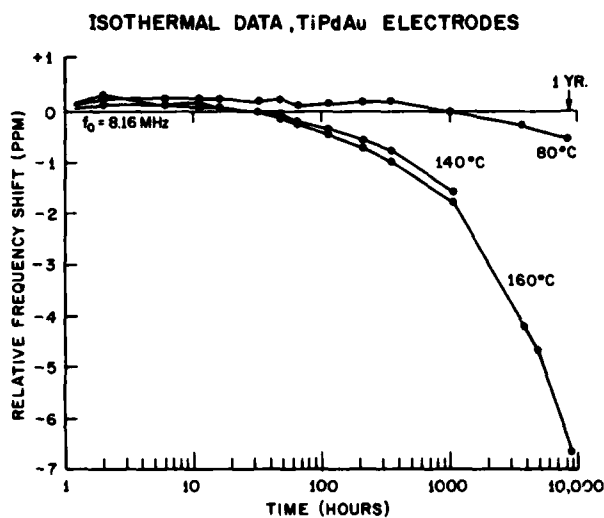


Figure 1

Complex isothermal aging curves for Ti Pd Au units.

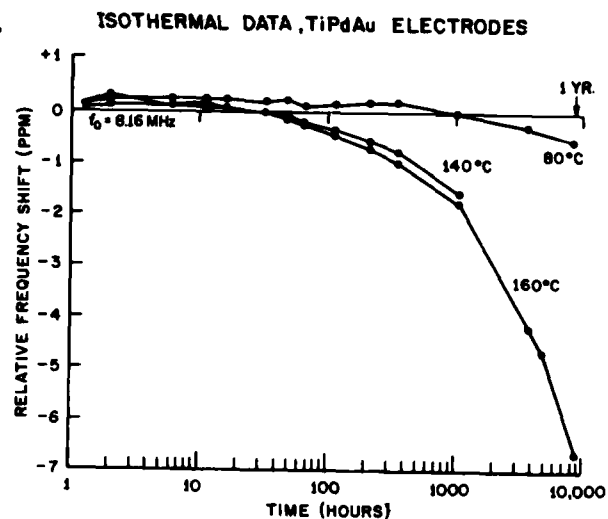


Figure 3

The accentuated curvature of the 160°C data is ideal for analysis.

AGING MODELS

1. $\Delta f = A_1 \ln t + A_2$
2. $\Delta f = A_1 \ln (1 + t/A_2) - A_3 \ln (1 + t/A_4)$
3. $\Delta f = A_1 (1 - e^{-t/A_2}) - A_3 (1 - e^{-t/A_4})$
4. $\Delta f = A_1 (1 - e^{-t/A_2}) - A_3 t^{A_4}$
5. $\Delta f = A_1 (1 - e^{-t/A_2}) - A_3 (t - A_4)^{0.5}$

t = TIME (HOURS)

A_1, A_2, A_3, A_4 = ADJUSTABLE COEFFICIENTS

Figure 2 The five aging models.

FIT QUALITY FOR THE FIVE MODELS (160°C)

MODEL	160°C DATA
	SUM OF THE SQUARES OF THE RESIDUALS (Hz ²)
1. $\ln t$	1892.5
2. DOUBLE \ln	30.0
3. CAPACITOR + CAP.	66.7
4. CAP. + t^{A_4}	10.6
5. CAP. + $t^{0.5}$	16.0

BEST
FIT
MODEL

Figure 4 Fit quality for the five models (160°C data).

BEST COMPUTER FIT (160°C DATA)

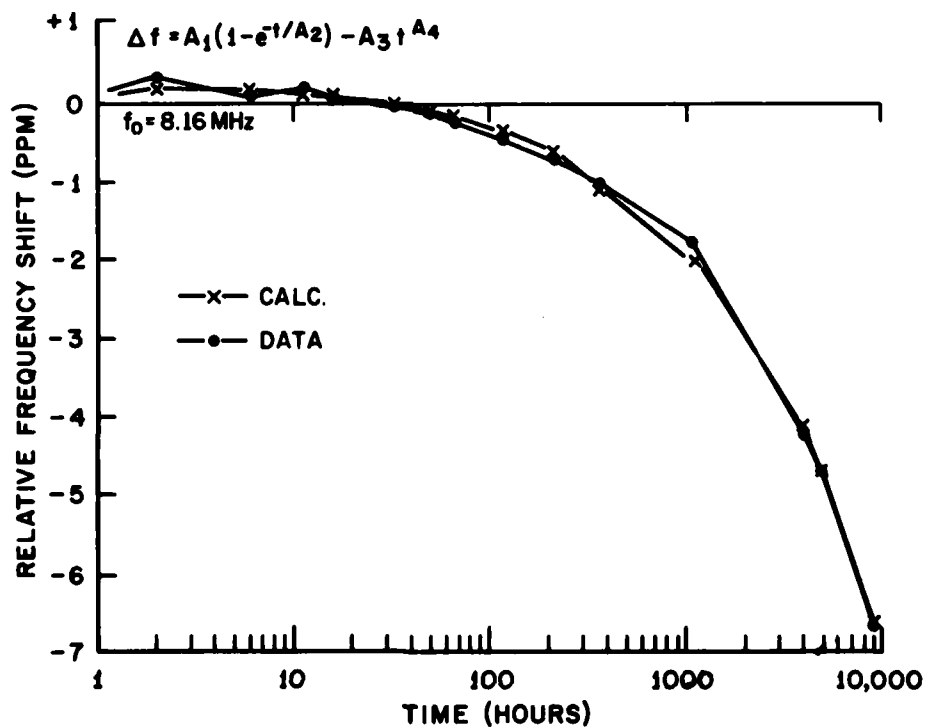


Figure 5 Best computer fit for the 160°C data.

BEST FIT A_4 COEFFICIENT VALUES

TEMPERATURE	A_4 IN t^{A_4} MODEL
80 °C	0.635
140 °C	0.492
160 °C	0.533
AVERAGE 0.553	

Figure 6
The best fit A_4 (Equation 4) coefficient values.

PROPOSED TiPdAu DIFFUSION AGING MODEL

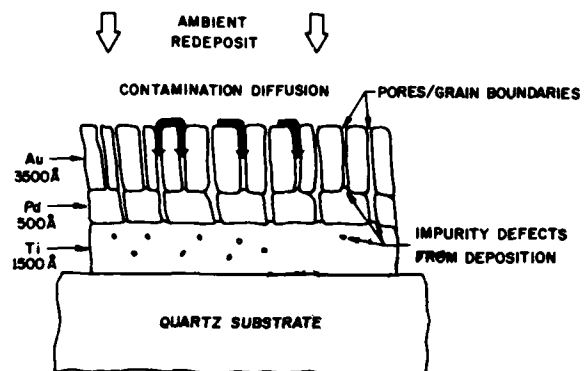


Figure 7
Schematic drawing of the proposed Ti Pd Au diffusion aging model.

THE ANALYSIS LOOP IN AN ASSEMBLY PROCESS

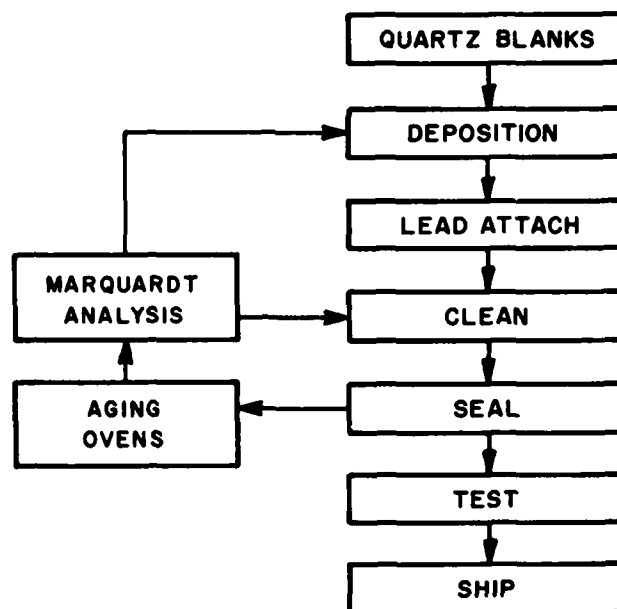


Figure 8
The analysis loop in an assembly process.

A NEW "ELECTRODELESS" RESONATOR DESIGN ⁺

Raymond J. BESSON

Ecole Nationale Supérieure de Chronométrie et de Micromécanique
La Bouloie - Route de Gray
25030 BESANCON CEDEX, FRANCE

Summary

In a recent paper, new structures, all using uncoated crystals, have been out-lined (and called B.V.A._n designs).

The main purpose of this paper is to give a complete description of the B.V.A.₂ design. In this design, the active part of the crystal is connected to the dormant part by little quartz "bridges", which can be very precisely made and located. As the fixation is made out of quartz, there is no ordinary bonding and among the consequences no discontinuity is caused by the fixation. Also, it is possible to avoid the stresses which could be caused by the machining of the quartz "bridges". Since the electrodes are located on insulators very close to the active surface of the wafer, the frequency of the device can be easily adjusted by means of a serie capacitance. Nevertheless, the construction of the device allows a very accurate frequency adjustment. As consequences of the construction acoustic losses are extremely low, a very high Q factor is obtained and the short term and long term stabilities are improved.

The first part of the paper deals with the construction parameters (wafer evaluation, excitation conditions and resonator mounting). Especially, the shape, dimensions and location of the quartz "bridges" are studied. The influence of the gap is discussed.

In the second part of the paper, resonators constructed according to the design are studied by various techniques. Acceleration effects are discussed. For comparison purposes, numerical data concerning AT 5 MHz fifth overtone units are given (electrical parameters, frequency spectrum, stabilities, amplitude frequency effect,...). Nevertheless, some results concerning other frequencies and S.C. cuts are also given. The problem of industrial fabrication is quickly discussed.

As a conclusion, the various results are reviewed and discussed (from the fundamental and technical points of view as well).

Key Words : Piezoelectric Resonator, Quartz Unplated Crystals, Electrode Effects, Bonding Effects, Frequency Stability, Aging, Frequency Adjusting, Crystal's Noise Reduction.

Introduction

In a recent paper ¹, we pointed out that mastering the boundary phenomena could reduce the aging and crystal's noise contribution as well. New structures, all using uncoated crystals, were outlined and called B.V.A._n designs :

- if *n* is odd, a rather conventional bonding and a special fixation is used. (In the B.V.A.₁ type, described last year, the main feature is that the crystal is "electrodeless" and frequency modifiable).

- if *n* is even, the design uses improved bonding and mounting.

This denomination indicates two successive steps of our attempt to reduce the crystal's noise and frequency drift contribution.

Since the B.V.A.₃ design is an improvement of the B.V.A.₁ design it will be very rapidly described using the scheme of Fig.1. The vibrating quartz crystal C of a given cut, orientation, geometrical shape (in Fig.1 a planoconvex disk) is, for instance, TC bonded (3 or even 4 bonding points ref.T) to the lower disk D₁ (which has been given a curvature identical to the curvature of the wafer's lower surface). D₁ is usually made out of quartz of the same cut and orientation. The electrodes are evaporated on the lower disk D₁ and the upper disk D₂₁ or D₂₂. The upper disk is not necessarily made out of quartz and may have any radius of curvature. The intermediate ring R determines the upper gap giving access to frequency adjustment or modulation. Compared to B.V.A.₁ type this design has mainly the same properties but the characteristic features are greatly improved in severe environmental conditions.

B.V.A.₂ RESONATORS

Special emphasis is given to this design which overcomes some difficult problems caused by the conventional evaluation of piezoelectric resonators. We mainly describe here quartz material 5 MHz units but, of course, other piezoelectric materials can be used and resonators of various frequencies have been evaluated.

⁺This work is supported by the "Direction des Recherches et Moyens d'Essais", D.R.M.E.

1 - Introduction and general description :⁺

Our goal was to obtain an "electrodeless" resonator (so as to overcome the difficulties due to electrode deposition) with a fixation exhibiting neither discontinuity nor local stress in the fixation areas. We wanted to obtain a device the frequency of which could easily be adjusted by means of a series capacitance. Then a large gap capacitance is suitable i.e. electrodes have to be located very close to the active surface of the wafer (in the micron range or even the 10 microns range). Also, and this is very important too, we planned to obtain fixation areas very accurately located and fixation means precisely known.

The E.V.A. resonator is represented by the schemes of Fig.2 - Fig.3 - Fig.3a - Fig.6 and Fig.7 and the pictures of Fig.4 and Fig.5. It includes :

- a vibrating quartz crystal, ref.C, the surface of which has been very carefully prepared. The active part of the crystal is connected to the dormant part by little quartz "bridges" very precisely made and located.
- a quartz condenser made of two disks (ref. D₁ and D₂) of the same cut and orientation on which the electrodes are deposited.
- means to maintain the condenser and crystal tightened together (it can be those recently described¹).
- a metallic experimental enclosure which is sealed by a pinch off process (a special coldwelded type enclosure has been made but not tested yet).

It must be pointed out that some construction parameters, especially the support configuration parameters, can be, using this design, very precisely known. Also since the crystal is "electrodeless" and uses an all quartz structure it is very suitable for low temperature applications². Moreover, the electrodes may be deposited on insulators which have been given a curvature different from the crystal surface's. This feature gives access to additional possibilities and may be used to modify Q factor, motional parameters series resistance and frequency amplitude effect.

Such a resonator, being entirely different from a conventional resonator, needs theoretical and technical studies specially devoted to it.

2 - Evaluation of the vibrating crystal :

The original part of this evaluation will only be described here. By use of ultrasonic machining and precise lapping³ little bridges are left between the external dormant part of the crystal and the internal vibrating part. Those bridges have a given shape, a given thickness, a given length. The schemes of Fig.6 outlines various possible shapes. The bridges can be very precisely located with respect to the thickness of the crystal (accuracy of the location : $\pm 10\mu$). Their angular position can also be very precisely known ($\pm 0.04^\circ$). Of course, the technique has to be perfectly mastered (for instance, avoiding a conical ultrasonic machining is not immediate) but, with sufficient experience, the

process can be considered as sure, rapid (2 or 3 minutes) and very accurate. As a consequence the middle part of the bridges can be located at the very nodes of vibration. Also, unwanted modes can be better eliminated. Since the thickness in the middle part of the bridges has ranged from 50μ to 1200μ (the usual is approximately 200μ) the bridges are not especially brittle. Any number of bridges can be left. Especially one single bridge, covering 360° angularly, may be directly lapped so avoiding the ultrasonic machining (Fig.3 a).

It must be pointed out that the machining does not destroy the material from the crystallographic point of view as can be seen from Fig.8 (vibrating 5 MHz fifth overtone, SID or Lang topography).

Moreover, no additional stresses are left by the machining if the quartz wafer is subjected, prior to mounting, to annealing at about 480°C , followed by a very slight surface attack with bifluoride.

The length and thickness of the bridges have been theoretically⁴ studied. Assuming a flexure vibration of the bridge, it is found that a length of 2 mm and a thickness of 0.2 mm is a good compromise between a weak static strain and a minimum acoustical energy transmission between the vibrating and dormant part of the crystal (5 MHz fifth overtone).

3 - Reflection of the elastic waves and influence of the gap :

The reflection of the elastic waves is not influenced by the position of the electrodes with regard to the crystal surface. It mainly depends on phenomena which occur in the boundary neighbourhood and which are due to crystalline modifications caused by machining processes and surface preparation.

The sample surface is carefully lapped and polished, so as to reduce the layer in which acoustic dissipation occurs. Defects due to machining processes are carefully investigated (X ray topography, electron microscopy and so on) so as to define the best procedures⁶. As far as possible, we operate in a clean room atmosphere, try to process the crystal in dry nitrogen and, of course, use the results of recent investigations for cleaning and decontamination⁶.

The influence of the gap has been studied. Experimentally, the Q factor is not a constant versus the total gap³. The variation depends on the frequency and overtone number of the unit. Of course, the variation is not important (smaller than 10 p.cent for gap variations from 0 to 1 mm). Nevertheless an investigation was started and proved that usual equivalent circuit is not sufficient⁵. So, starting from the exact expression of the current, we computed the Q factor versus the gap (assuming a plane infinite plate) and found a variation which gives a better account of the results.

Actually, a compromise must be chosen. The series resistance and the motional inductance strongly increase with the gap. Also, the frequency of the unit must be easily adjusted by a series capacitance ; so very thin gaps are suitable.

⁺F. Patent 76 01 035 - 76 16 289.

But the mechanical stability of the gap thickness is to be considered too, if ultrastable units are desired. For a 5 MHz fifth overtone we use gaps in the micron or 10 microns range. Nevertheless for resonators on the fundamental mode the gap can be larger.

Usually, the gaps are made by a special lapping³ process which affects the central area of D_1 and D_2 (see Fig.2). They can also be made by nickel electrodeposition as suggested by Fig.7. It must be pointed out that slightly different gaps can be made so giving access to very precise frequency adjustment (1 Hz for a 5 MHz fifth overtone unit).

Main Characteristics of B.V.A.₂ Resonators

As pointed out last year, B.V.A.₂ resonators are given more interest in the high frequency range since the electrode and bonding phenomena are relatively larger for high frequency crystals. Nevertheless, for comparison purposes, numerical data concerning AT 5 MHz fifth overtone units are given.

1 - Electrical parameters :

Resonators evaluated with good natural quartz correspond to Q factor and series resistance given by :

$$Q = 3 \cdot 10^6 \quad R_1 = 60\Omega$$

We are presently conducting an investigation covering the following types of material :

- Natural Quartz (various origins)
- Electronic Grade Quartz (various origins)
- Optical Grade Sawyer Quartz
- Premium Q Sawyer Quartz
- Premium Q Sawyer Swept Quartz.

2 - Stabilities :

The results concerning the short term stabilities have already been given¹ ; some improvements have been obtained but they are not significant.

Long term drift experiments have been carried out in our Laboratory (M. Decailliot and J. Chauvin). The reference is a Cesium beam standard. A Butler oscillator is used ; the level of oscillation, which is about 1 μ W, is regulated (approximately to 10^{-4}) and the temperature is stable to better than 10^{-3} °C over large periods of time. The signals are 10 times frequency divided ; then phase compared. The intervals of time between two phase coincidences are automatically recorded. Using B.V.A.₂ crystals, not pre-aged at all, the following drifts per day have been obtained :

- immediately : 2 to $2.5 \cdot 10^{-10}$
- after a month continuous operation : $3 \cdot 10^{-11}$
- after two months continuous operation : $7 \cdot 10^{-12}$ to $1 \cdot 10^{-11}$

The experiments are going on but the regular decrease of the drift is to be pointed out.

Nevertheless, this result must not be regarded as definitive as long as we don't use a coldwelded type enclosure. (in our experimental soldered enclosure the crystal is contaminated and it cannot be baked and pumped down at temperatures higher than the solder melting point).

3 - Static acceleration effects :

B.V.A.₂ resonators of various types (different support configuration, plano-convex or bi-convex units) have been studied and tested in the "Office National d'Etudes et de Recherches Aéronautiques" (ONERA) by M. Valdois and D. Janiaud.

The experiments performed are principally related to the influence on the resonator frequency of the direction of an acceleration of constant modulus. Resonators have been tested either in passive networks or in an oscillator loop. Acceleration vectors in three orthogonal planes associated to the resonator are applied. Experimental results are similar to those obtained with customary 2 or 3 support units. All frequency deviations are sinusoidal functions of azimuth angle (the zeros of these functions define null influence directions). Those 3 directions of null influence determined by studying the 3 associated planes, belong to a same plane which is called "accelerometric null influence plane"⁷. Thus, it is proved that a plane exists in which any acceleration has no measurable influence on the resonator frequency (the frequency variation $\frac{\Delta f}{f}$ is measured with a 10^{-11} accuracy). As a consequence the direction perpendicular to this plane is the maximum influence direction.

Mainly, the following B.V.A.₂ AT resonators have been tested :

- plano-convex resonator with a single 360° bridge. Maximal values of the sensitivity from $4 \cdot 10^{-9}$ /g to $8 \cdot 10^{-9}$ /g have been recorded. It must be pointed out that the symmetry of the bridge with respect to the center of the resonator was not guaranteed by the machining.
- plano-convex four bridge resonator (bridges along Z Z' and X X'). Maximal sensitivities of $1.5 \cdot 10^{-9}$ /g to $2 \cdot 10^{-9}$ /g have been recorded. (This is slightly less than values recorded for a traditional two or three support unit).
- bi-convex ($R_1 = R_2 = 150$ mm) four bridge resonator. The maximal sensitivity is lower than 10^{-9} /g.

In any case, no residual frequency deviation is observed when the acceleration is suppressed. Recent theoretical studies have shown the influence of support configuration⁸ or slight dissymmetries⁹ on accelerometric sensitivity. Application of these results to B.V.A.₂ resonators appear as rather complex because of the great variety and special features of the support configurations. Experimental results confirm the great influence of the fixation and the interest of bi-convex contours.

4 - Other cuts and other frequencies :

B.V.A.₂ resonators have been made using other cuts especially S C cuts. (5 MHz fifth and third overtones 10 MHz third overtone). Interesting results have been obtained (especially much lower amplitude frequency effect).

The acceleration sensitivity is found to depend largely on some parameters and especially on the fixation configuration and evaluation (we preferentially use a four bridges support configuration).

Resonators 100 MHz fifth overtone have also been constructed and encouraging results have been obtained (high Q factors and easy frequency adjustment).

5 - Evolution and cost of the fabrication processes :

Over one hundred crystals have been evaluated, generally unit by unit. When greater quantities are made some operations can be simplified (especially frequency adjustment). To our experience, it appears that the fabrication cost of a B.V.A.₂ unit can be reduced to 1.2 or 1.5 time the cost of a traditional unit of the same frequency.

Conclusions

By use of a fixation made out of quartz the discontinuities and the stresses, usually caused by the traditional bonding processes, are avoided. The fixation location and characteristics are very precisely known. Since the electrodes are located very close to the active surface of the vibrating crystal, the frequency of the device can be easily adjusted. Nevertheless the gaps evaluation allows a very accurate frequency adjustment.

As consequences, very low acoustic losses and improved stabilities are obtained. It also appears that some properties related to acceleration effects are interesting (especially there is no permanent residual frequency deviation).

B.V.A.₂ resonators are interesting for fundamental studies, since their structure is very different from the usual. (precision of fixation means, properties related to heat transfer and crystal's noise contribution, additional construction parameters,...).

Those provisional results can provide direction for future research in the field of uncoated resonators and indicate promissive developments.

References

- 1 - R. Besson Proc. 30th A.F.C.S. p.p. 78 - 83 1976
- 2 - AG Smagin Pribery i Teckhnika Eksperimenta N°6 p.p. 143 - 145 Nov. Dec. 1974
- 3 - P. Maitre E.N.S.C.M.B. (part of a thesis work)
- 4 - B. Dulmet E.N.S.C.M.B. and J. Piranda - Laboratoire de Mécanique Appliquée - Besançon.
- 5 - D. Gillet B. Dulmet E.N.S.C.M.B. (part of a thesis work).
- 6 - J.R. Vig J.W. Lebus Proc 29th A.F.C.S. p.p. 220 - 229 1975
- J.R. Vig, C.F. Cook, Jr, K. Schwidtal, J.W. Lebus, E. Hafner Proc. 28th A.F.C.S. p.p. 96-108 1974

7 - M. Valdois - D. Janiaud to appear in Applied Physics letters.

8 - D. Janiaud C R Acad. Sc. Paris.

9 - P C Y Lee and Kuang-Ming-Wu Proc. 30th A.F.C.S. p.p. 1 - 7 1976.

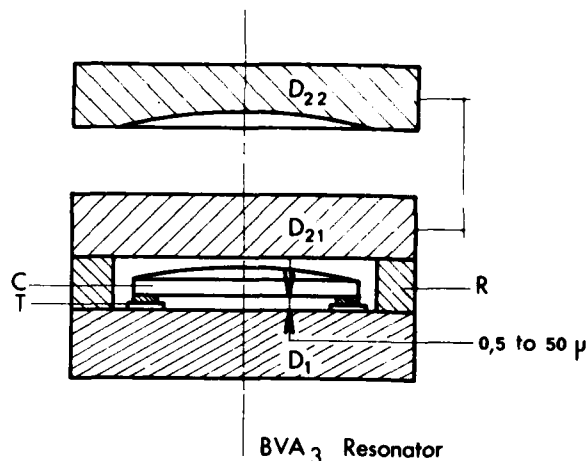


Figure 1. BVA₃ Resonator

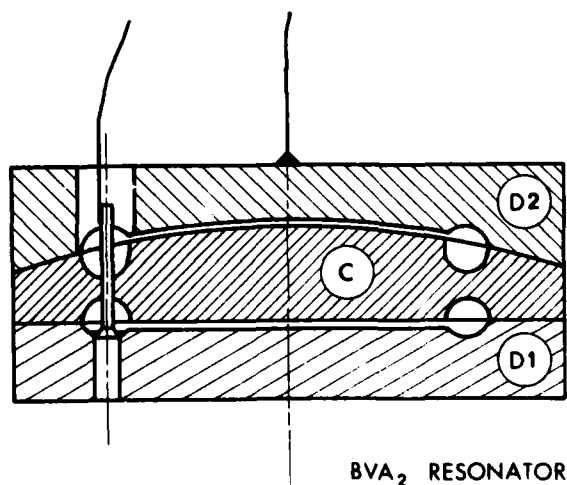


Figure 2. BVA₂ Resonator

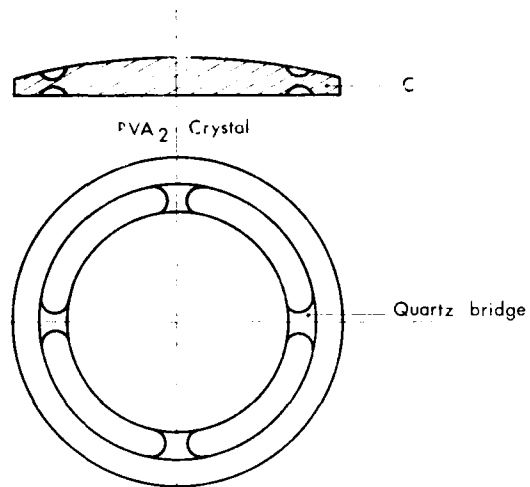


Figure 3. BVA₂ Crystal

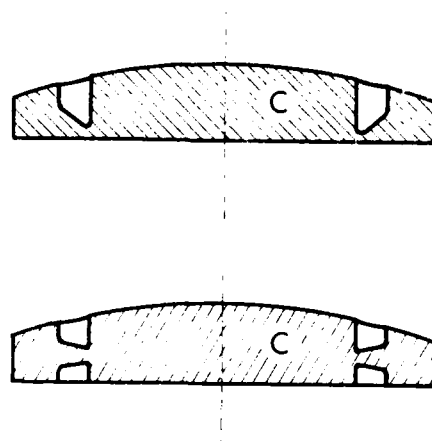
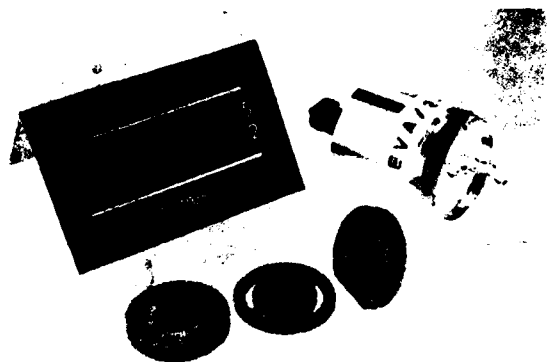
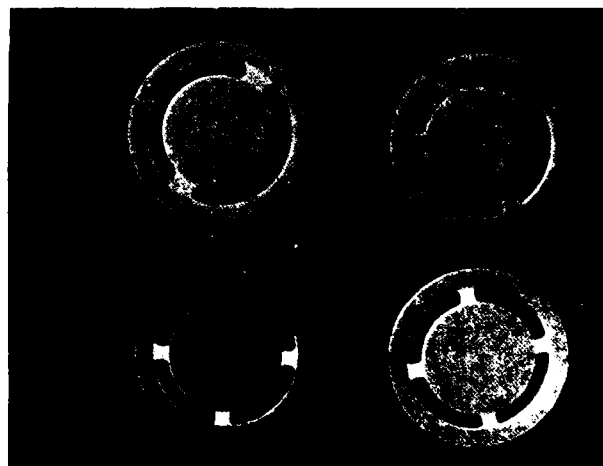


Figure 3a. BVA₂ Crystal - Single Bridge



BVA₂ RESONATOR

Figure 5. BVA₂ Resonator



BVA₂ CRYSTAL

Figure 4. BVA₂ Crystal

BRIDGES - VARIOUS SHAPES

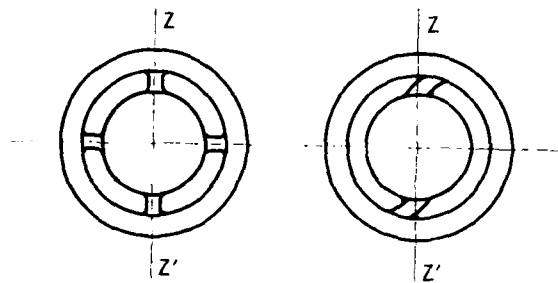
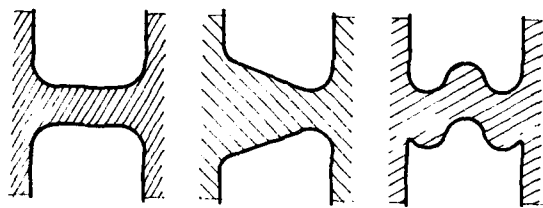


Figure 6. Bridges - Various Shapes

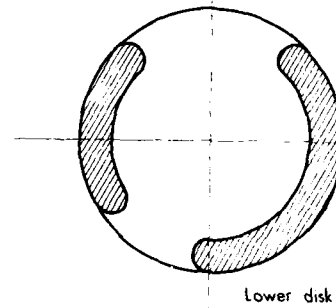
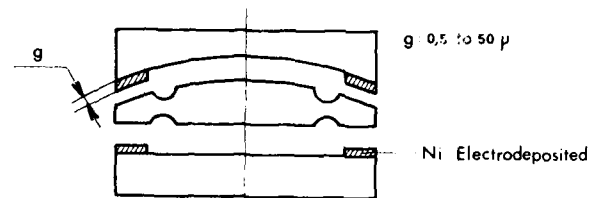


Figure 7. BVA_2 Resonator

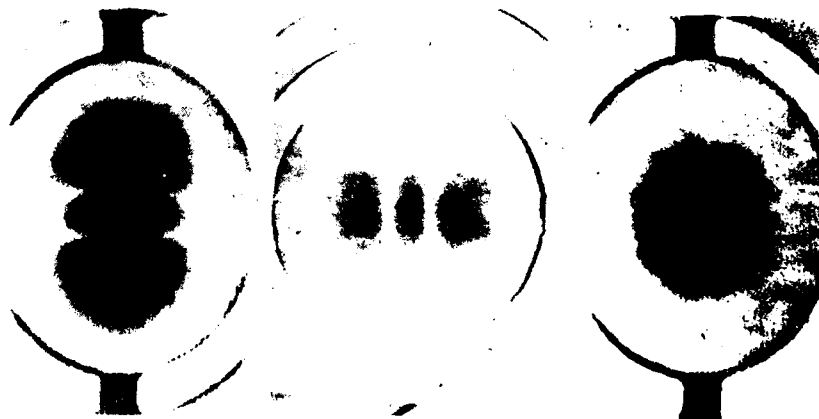


Figure 8. Electrodeposition patterns

MAKING DOUBLY ROTATED QUARTZ PLATES

Walter L. Bond
Stanford University
Stanford, California

John A. Kusters
Hewlett-Packard Laboratories
Palo Alto, California

Introduction

Although the AT cut quartz plate has a zero temperature coefficient of frequency at a certain temperature, it is sensitive to sudden small temperature changes. It is also sensitive to forces of acceleration.

The AT plate is a 'singly rotated' orientation. In IEEE nomenclature,¹ it is specified as a $(yx\lambda)\theta$ cut. This means that the plate is related to an unrotated reference plate that has its thickness in the Y direction of the crystal, and its length in the X direction (see Fig. 1). The rotated cut is considered as being the reference crystal rotated counterclockwise by amount θ about its length, as shown in Fig. 2. For an AT plate θ has the value 35.25° , i.e., it is $(yx\lambda)35.25^\circ$ cut. There are several other zero temperature coefficient cuts with different modes of vibration, all developed before 1941; such as the BT, $\theta = -49^\circ$, CT, $\theta = 38^\circ$, DT, $\theta = -52^\circ$, ET, $\theta = 66^\circ$, and the FT, $\theta = -57^\circ$. There are other orientations for which it is easier to start with a different reference orientation. One such is the $(xyt)-18^\circ$, many of which were made.

The planning of these orientations depended on measurements of the elastic constants and expansion coefficients, assuming these to be linear. However, such relations are never strictly linear. Quartz has six elastic constants relating stress to strain. Measurements of higher order terms have been made and orientations planned that greatly improve performance of such oscillator plates.²⁻⁵ These orientations require a second rotation from the reference position, as shown in Fig. 3.

Doubly Rotated Plates

The making of doubly oriented cuts can be confusing, because one must keep track of several things, any of which if done wrong will result in a greatly inferior product.

Preparation

We will assume that we wish to make $(yx\omega\lambda)\phi/\theta$ plates and that we will slice them off the end of a bar, not straight across, but at a complex angle which we must determine. We will make the bar of convenient length (along the Y axis) and convenient cross-section. Two opposite long faces of the bar are perpendicular to the optic axis, that is these faces are (00.1) planes. The other two long faces are perpendicular to the X axis, that is they are (11.0) planes. We must determine which of these is to be taken as the positive X face. The X and Z faces may need to be X-ray corrected.

Quartz comes in two kinds, right-handed and left-handed. A simple optical test distinguishes these. Looking along the optic axis with the sample between polarizing filters, one sees concentric rings about the optic axis. If these rings expand when the analyzer is turned clockwise, the quartz is right-handed; if they contract, the quartz is left-handed.⁶

For right-handed quartz, compression along the X-axis develops a positive electric charge on the +X face, a negative charge on the -X face. For left-handed quartz, compression along the X-axis develops a negative charge on the +X face, and a positive charge on the -X face.⁶ These charges can be detected by means of a sensitive voltmeter, or other suitable detectors.

Another method which determines the positive X face is visual inspection of etch pit patterns.⁷⁻⁸ After heavily etching the X faces in a suitable etchant such as ammonium bifluoride in a sugar solution, a pin-hole oriascope may be used to determine in one test whether the quartz is right- or left-handed, and which X face is the positive one.

We may also initially buy cultured quartz which has been marked by the vendor. Whatever method is used, we now stain the +X face of the bar red. Since the X axis is an axis of two-fold symmetry, the bar may be turned 180° about X if necessary, thus reversing Y and Z without affecting the behavior of the finished plates.

Slicing

A plate $(yx\omega\lambda)\phi/\theta$ has its thickness direction along a unit vector

$$\vec{t} = \begin{pmatrix} -\cos \theta \sin \phi \\ \cos \theta \cos \phi \\ \sin \theta \end{pmatrix}$$

as can be seen from Figs. 1, 2, and 3.

We define a set of 'saw coordinates', S_1 along the direction of the sliding saw table motion, S_2 parallel to the saw blade axis and S_3 to form an orthogonal right-handed unitary system, see Fig. 4. We place the crystal bar on the saw table, a Z face up, +X (red face) toward us as shown in Fig. 5.

If ϕ and θ are both positive, we now roll the bar about S_2 , the top going away from us. The angle of roll is β . This roll changes the component of \vec{t} on S to:

$$\begin{aligned} \vec{t} &= \begin{pmatrix} \cos \beta & 0 & -\sin \beta \\ 0 & 1 & 0 \\ \sin \beta & 0 & \cos \beta \end{pmatrix} \cdot \begin{pmatrix} -\cos \theta \sin \phi \\ \cos \theta \cos \phi \\ \sin \theta \end{pmatrix} \\ &= \begin{pmatrix} -\cos \beta \cos \theta \sin \phi - \sin \beta \sin \theta \\ \cos \theta \cos \phi \\ -\sin \beta \cos \theta \sin \phi + \cos \beta \sin \theta \end{pmatrix} \end{aligned} \quad (1)$$

If $\tan \beta = \tan \theta / \sin \phi$ the S_3 component vanishes. A rotation of amount γ clockwise about S_3 is given by:

$$\vec{t} = \begin{pmatrix} \cos \gamma \sin \gamma & 0 \\ -\sin \gamma \cos \gamma & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} -\cos \beta \cos \theta \sin \phi & -\sin \beta \sin \theta \\ \cos \theta \cos \phi & 0 \\ -\cos \gamma \cos \beta \cos \theta \sin \phi & -\cos \gamma \sin \beta \sin \theta \\ & + \sin \gamma \cos \theta \cos \phi \\ \sin \gamma \cos \beta \cos \theta \sin \phi & + \sin \gamma \sin \beta \sin \theta & + \cos \gamma \cos \theta \cos \phi \\ & & 0 \end{pmatrix} =$$

Since \vec{t} is unitary, if $t_1 = 0$, then $t_2 = \pm 1$. This occurs for

$$\tan \gamma = \cos \beta \tan \phi + \sin \beta \tan \theta / \cos \phi. \quad (2)$$

From Eqs. (1) and (2) we derive

$$\tan \gamma = \tan \phi / \cos \beta, \quad (3)$$

or

$$\cos \gamma = \cos \phi \cos \beta. \quad (4)$$

To mount this on the saw, we use matched wedges of angle $\delta = 90^\circ - \beta$ and use a protractor to position the bar at angle γ as shown in Fig. 6. Now dental plaster is pushed under and around the center of the bar. This hardens within 15 minutes. The wedges are then removed and more dental plaster is pushed under the ends. When this hardens, we may slice off the wafers.

This procedure can be greatly simplified if we can obtain cultered quartz where the first rotation about the width is already "grown" into the bar.⁹ Then if the bar axes are suitably identified, slicing crystal plates is the same as cutting any of the singly-rotated cuts such as the AT or the BT. Fig. 7 shows such a bar and several slices cut from it.

X-Ray Orientation

In general, there are no short-cuts in X-ray orientation of quartz crystals except that provided by natural symmetry. A doubly rotated crystal cut only further complicates the process.

Fig. 8 shows an orthographic projection of the primitive region of quartz. The location of the allowable diffraction planes using copper $K\alpha$ radiation are shown. For this, and for all subsequent X-ray work, we assume that the wavelength is that of copper $K\alpha_1$ at 18°C., namely 1.54056 Å. On this figure, all of the singly rotated common cuts and several of the doubly rotated cuts such as the IT, the RT, the FC, the LC, and the focus of this paper, the TTC, are shown. The IT was originally developed to take advantage of several of the properties of doubly rotated cuts and was located on the sequence of planes at about $\phi = 19^\circ$. X-ray orientation on a sequence of planes such as these allows relatively accurate determination of at least one direction.

The LC, the FC, and the TTC have no such advantage. Accurate determination of any of these orientations requires the use of at least two, and preferably three, non-coplanar atomic planes.

Fig. 9 is a diagrammatic representation of a typical X-ray diffraction. θ_B is the atomic plane Bragg angle given by:

$$N\lambda = 2d \sin \theta_B \quad (5)$$

where d is the atomic spacing and λ is the radiation wavelength. δ is the angle between the surface normal and the normal to the atomic plane, and A and B are as shown. A and B are easily calculated¹⁰, knowing the desired surface and the Miller indices of the plane. Doubly rotated cuts such as the TTC generally

require that we use a double-crystal goniometer¹¹, as shown in Fig. 10. Working to a desired turnover temperature for the TTC requires particular care in the crystal orientation. A 1°C tolerance on turnover temperature requires that the orientation be held to within 10" of arc. The standard rocking curve for a single-crystal goniometer setup is usually several minutes wide. The use of a reference crystal can reduce the width of the rocking curve sufficiently to allow a precision of about 6" of arc. Fig. 10 shows a situation where the Bragg angles of the two planes are different. This may be due to either using different planes, or using different orders of the same plane. In either case, a problem develops. The double-crystal system has sufficient resolution to fully separate the $K\alpha$ doublet. The characteristic $K\alpha$ lines for copper are shown in Fig. 11. As a result, the detector shows two separate and usually well-resolved, radiation peaks. Fortunately, the proper peak, the $K\alpha_1$, is the stronger of the two and always occurs at a smaller goniometer angle.

The double peak can be eliminated if the plane used for the test crystal and the reference crystal are identical. Then the two planes are exactly parallel at the proper orientation, and the double peak becomes a single narrow radiation peak.

Another complication is that the simple picture shown in Fig. 9 is valid for only one particular rotation of the test crystal. As the test crystal is rotated about its normal, the plot of goniometer reading vs. rotation angle is similar to that shown in Fig. 12. Ideally, we would like to orient the crystal in rotation such that we are always operating at the very bottom of this curve. At this point, any possible rotation error produces the minimum error in the goniometer reading.

TTC Orientation

To make crystals at the TTC orientation, (yxw1) 21.93°/34.11°, we might return to an orthographic projection similar to that shown in Fig. 8. A simpler version is given in Fig. 13. In general, we would like to use two planes that have essentially the same Bragg angle, and that are also about 90° apart in rotation, as seen at the test crystal. The first condition allows us to use two crystal planes with the same Bragg angle as a reference crystal. This produces a single radiation peak on the detector readout, simplifying the measurement. The second condition yields either a very rapid convergence, if the test crystal is being prepared in a special fixture for lapping to the desired surface, or the fastest determination of the present orientation of the test crystal surface using the methods outlined in the Appendix. The two planes that most closely meet these conditions are the (12·3) and the (21·3) as shown. The two planes are usually sufficient to determine the surface accurately if we keep careful track of the rotation of the test crystal. As a check, either of the two other planes shown, the (12·1) or the (12·2) are also useful. From our experience, the (12·1) is perhaps better assuming that we also keep in mind that this plane will always produce a double-peaked response with a (12·3) reference crystal.

The various angles, detector settings, etc. are shown in Table 1. The data for the (00·6) plane is given to provide a reference point for the crystal blank rotation angles.

Summary

Assuming a well-aligned reference crystal, the (12·3) and (21·3) planes can be used to orient the test crystal surface to within approximately 6" of arc.

A further check can be made using a (12·1) plane. Using the measured angles, the methods of the Appendix may be used to determine the actual surface orientation.

Memorial

While this paper was in preparation, Walter Lysander Bond died as a result of complications following heart surgery. After a career spanning forty years (1928-1968) at Bell Telephone Laboratories, Walter retired to become a research physicist at the W.W. Hansen Laboratories of Physics, Stanford University. It was my privilege to know Walter during this period when he regularly ran every grad student ragged trying to keep up with the wealth of ideas, experience, and engineering technology that he provided. When I first approached Walter with the idea of writing this paper, it was not envisioned to be a joint venture; his health dictated otherwise.

This paper, the last of his many efforts in crystal technology, was completed after his death. Any faults in it are due to my attempts to interpret the ideas we developed in the original working sessions several days before his heart surgery.

John A. Kusters

Appendix

The unit vector in the thickness direction of the test crystal was given previously as:

$$\vec{t} = \begin{pmatrix} -\sin\phi \cos\theta \\ \cos\phi \cos\theta \\ \sin\theta \end{pmatrix} \quad (A1)$$

Similarly,¹⁰ we can describe a unit vector that is normal to the desired atomic plane (hk·l) as:

$$\vec{n} = \frac{1}{S} \begin{pmatrix} h \\ \frac{h+2k}{\sqrt{3}} \\ \frac{l}{c/a} \end{pmatrix} \quad (A2)$$

$$\text{where } S = \left[h^2 + \frac{1}{3} (h+2k)^2 + \frac{l^2}{(c/a)^2} \right]^{\frac{1}{2}}$$

is the normalizing function, a and c are the atomic lattice parameters, and h, k, and l are the Miller-Bravais indices.

The angle between the desired surface and the atomic plane used to measure the surface is the same as the angle between the two normal vectors, or,

$$\cos \delta = \vec{t} \cdot \vec{n} \quad (A3)$$

In practice, we do not measure δ directly, but either angles A or B. The angle A is somewhat easier to measure accurately. Therefore, restricting ourselves to A

$$A = \theta_B - \delta \quad (A4)$$

$$\text{where } \theta_B = \sin^{-1} \left(\frac{N\lambda}{2d} \right) \quad (A5)$$

where d is the interatomic lattice spacing and is given by:

$$d = \frac{a}{S} \quad (A6)$$

From (A3), we have;

$$\cos \delta = n_1 t_1 + n_2 t_2 + n_3 t_3 \quad (A7)$$

where the subscripts refer to the i^{th} component of the unit vector.

Three Plane Solution

By measuring the goniometer angle for each of three known atomic planes, we have a set of three simultaneous equations in the three unknowns, t_1 , t_2 and t_3 .

$$\begin{pmatrix} n_{11} & n_{12} & n_{13} \\ n_{21} & n_{22} & n_{23} \\ n_{31} & n_{32} & n_{33} \end{pmatrix} \cdot \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} = \begin{pmatrix} \cos \delta_1 \\ \cos \delta_2 \\ \cos \delta_3 \end{pmatrix} \quad (A8)$$

where the n_{ij} 's are solely a function of h, k, and l. Any of several mathematical techniques may be used to solve this set. The solution yields the three components in (A1).

$$\theta = \sin^{-1} t_3$$

$$\phi = \cos^{-1} \left(\frac{t_2}{\cos\theta} \right) \quad \text{or} \quad \sin^{-1} \left(\frac{-t_1}{\cos\theta} \right)$$

and as a check on the 'goodness' of the solution,

$$t_1^2 + t_2^2 + t_3^2 = r^2$$

where for a perfect fit, $r = 1$.

Two Plane Solution

An iterative solution can be derived which requires the determination of only 2 planes. In this case:

$$-\sin\phi \cos\theta n_{i1} + \cos\phi \cos\theta n_{i2} + \sin\theta n_{i3} = \cos \delta_i$$

where $i = 1, 2$.

By proper choice of θ_0 , the starting point, we can iterate solutions on (ϕ, θ_i) such that the solution becomes stationary.

$$\text{i.e. take } \begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix} \cdot \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = \begin{pmatrix} \cos \delta_1 - \sin\theta_i n_{13} \\ \cos \delta_2 - \sin\theta_i n_{23} \end{pmatrix} \quad (A9)$$

$$\text{where } t_1^2 + t_2^2 = \cos^2 \theta_{i+1}$$

$$\frac{t_1}{t_2} = -\tan \phi$$

then continue solution until $\theta_i = \theta_{i+1}$ to within an arbitrary error tolerance. Because of the ambiguity of using 2 planes, 2 solutions are possible. One must set up the problem with this possibility in mind.

References

- ¹ ANSI C83.3-1951 (R1972), IEEE Standard 176-1949 "Standards on Piezoelectric Crystals," Proc. IRE, 37, pp 1378-1395, Dec. 1949.
- ² R. Holland, "Nonuniformly Heated Anisotropic Plates: II, Frequency Transients in AT and BT Quartz Plates," 1974 Ultrasonics Symposium Proceedings, pp 593-598.
- ³ E. EerNisse, "Quartz Resonator Frequency Shifts Arising from Electrode Stress," Proc. 29th Ann. Symp. on Frequency Control, pp 1-4, May 1975
- ⁴ J. Kusters, "Transient Thermal Compensation for Quartz Resonators," IEEE Trans. Sonics and Ultrasonics, SU-23, pp 273-276, July 1976.
- ⁵ J. Kusters and J. Leach, "Further Experimental Data on Stress and Thermal Gradient Compensated Crystals," Proc. IEEE, 65, pp 282-284, Feb. 1977.
- ⁶ Ref. 1, pg. 1386
- ⁷ R. Heising, Editor, Quartz Crystals for Electrical Circuits, Van Nostrand Co., New York, 1946.
- ⁸ W. L. Bond, Crystal Technology, Wiley, New York, 1976.
- ⁹ Sample supplied to us by Sawyer Research Products, Inc., Eastlake, Ohio.
- ¹⁰ Ref. 7, esp. Ch. 3, W. L. Bond and E. Armstrong, "The Use of X-Rays for Determining the Orientation of Quartz Crystals," pp. 95-139.
- ¹¹ W. L. Bond, "A Double-Crystal X-Ray Goniometer for Accurate Orientation Determination," Proc. IRE, 38pp 886-889, Aug. 1950.

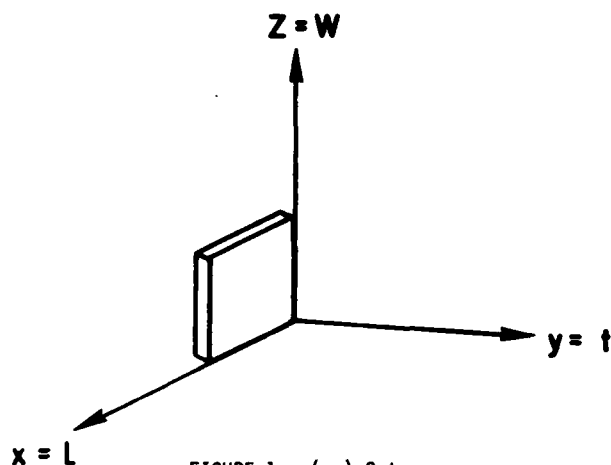


FIGURE 1 - (yx) Cut.

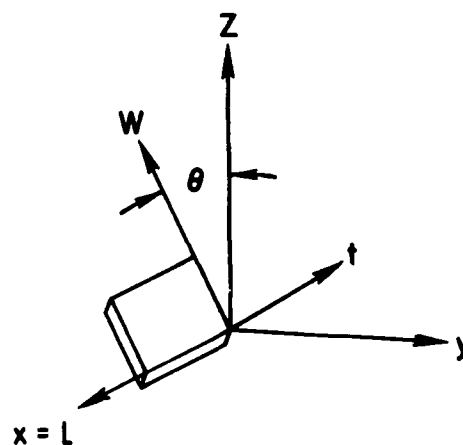


FIGURE 2 - (yxl)θ Cut.

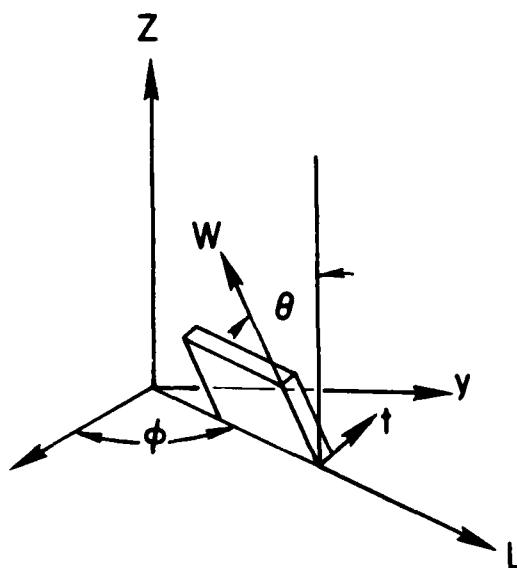


FIGURE 3 - (yxwL)φ/θ Cut.

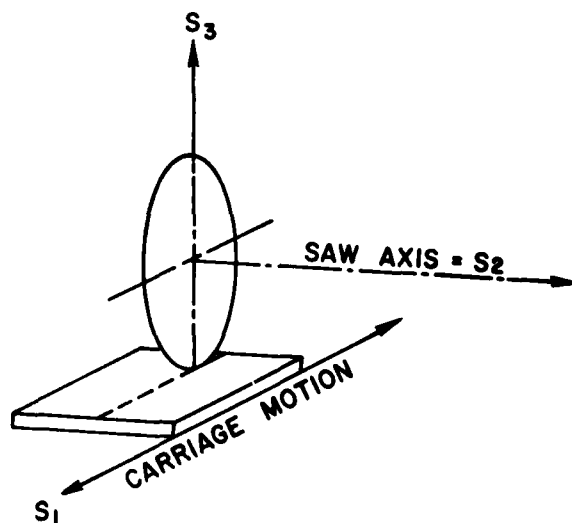


FIGURE 4 - Definition of 'SAW' Coordinate System.

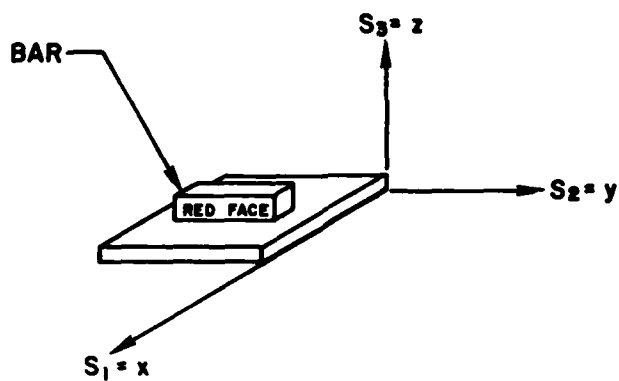


FIGURE 5 - Initial Quartz Bar Orientation.

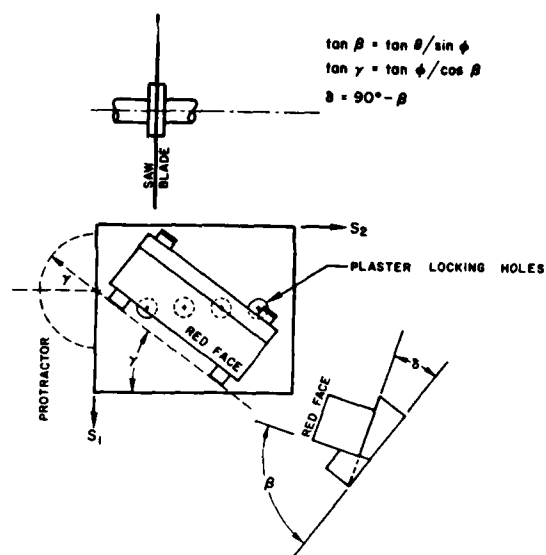


FIGURE 6 - Final Orientation, Ready for Cutting.



FIGURE 7 - Cultured Quartz 22° Rotated Bar.

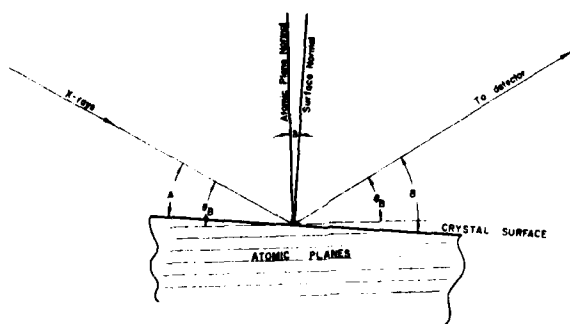


FIGURE 9 - Simple X-Ray Diffraction.

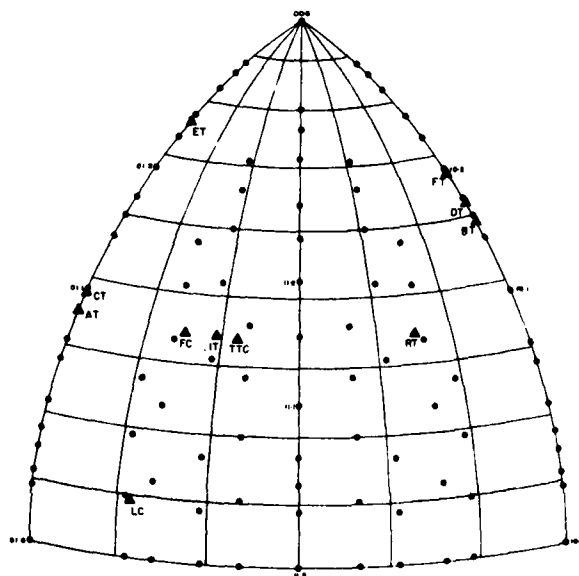


FIGURE 8 - Orthographic Projection of Primitive Region in Quartz Showing All Allowable X-ray Diffraction Planes.

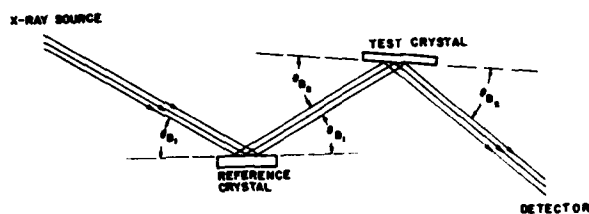


FIGURE 10 - Double-Crystal Goniometer Set-up.

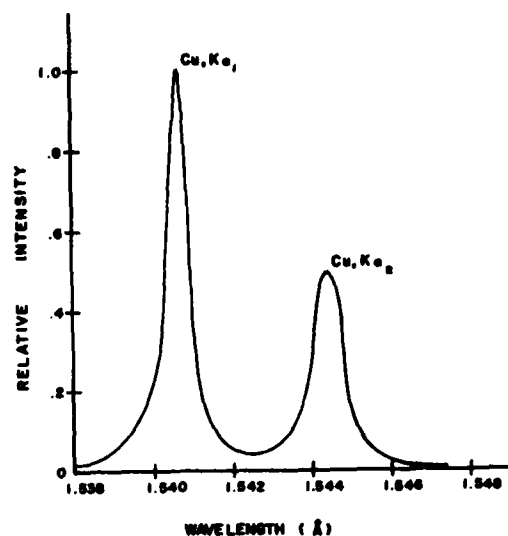


FIGURE 11 - Characteristic Radiation - Copper Target X-ray Source.

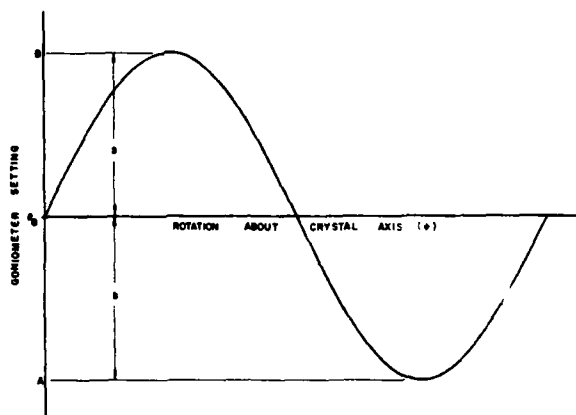


FIGURE 12 - Goniometer Position vs. Crystal Rotation.

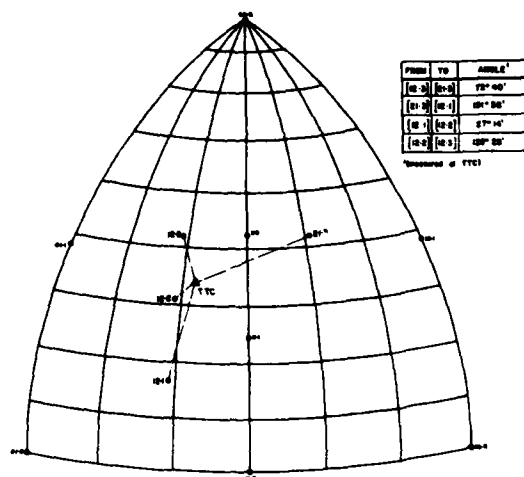


FIGURE 13 - Orthographic Projection Showing TTC and Recommended Planes.

ATOMIC PLANE	DETECTOR SETTING ($2\theta_B$)	ROTATION FROM Z-Z' AXIS	A	B
00-6	117° 32' 58"	0°	2° 53' 5"	114° 39' 53"
12-1	59° 57' 59"	189°	12° 15' 40"	47° 42' 18"
12-2	67° 48' 04"	216°	29° 48' 43"	37° 59' 21"
12-3	79° 53' 31"	348°	31° 59' 10"	47° 54' 21"
21-3	79° 53' 31"	57°	23° 12' 00"	56° 41' 31"

TABLE I - X-Ray Settings for (yxwl) 21.93°/34.11°.

THE ELECTROELASTIC TENSOR AND OTHER SECOND-ORDER PHENOMENA
IN QUASILINEAR INTERPRETATION OF THE POLARIZING EFFECT
WITH THICKNESS VIBRATIONS OF α -QUARTZ PLATES

Carl K. Hruska
Department of Mathematics, York University
Toronto, Ontario M3J 1P3

Summary

The change in series resonance frequency of piezoelectric resonators caused by a dc electric field superposed on the driving field is studied on the three thickness modes of thin α -quartz plates. For resonators operating in the fundamental frequency range between 2.7 and 5.5 MHz and at the dc electric field intensities between $\pm 10^6 \text{ Vm}^{-1}$, the change in frequency is proportional to the intensity of the applied dc field and depends on the crystallographic orientation of the plates and on the mode of vibration. At the dc field of 10^6 Vm^{-1} , the maximum change in frequency is found to be 16 parts per million.

Two sets of experimental data are presented: 32 values for the cuts of orientation (xzl) $\frac{1}{2}$ and 29 values for the cuts of orientation (xzw) $\frac{1}{2}$.

Quasilinear interpretation relates the observed frequency changes to the converse piezoelectric effect and to the tensors of electroelasticity, second-order piezoelectricity, and second-order permittivity. Several components of these three tensors are calculated.

An analysis of the results indicates that the change in frequency is caused mainly by the change in the values of the elastic constants. The contributions of the second-order piezoelectricity and the second-order permittivity are of lesser importance. The available data on the frequency change can also be interpreted by considering only the electroelastic and the converse piezoelectric effects. Several components of the electroelastic tensor are calculated under these simplified conditions. Good agreement is obtained with the earlier results by Kusters.

In general, the ability of the calculated second-order tensor components to reproduce the experimental data is good, the differences being several percent on the average.

Considering the quasilinear character of the approach adopted in this work, an allowance must be made for the cumulative nature of the second-order constants calculated.

The analyzed phenomenon may well become a convenient and sensitive tool for the study of a variety of higher-order effects in α -quartz and other materials.

Key Words : Piezoelectric resonator, frequency shift, electric field effect, quartz, second-order effects, electroelastic tensor, second-order piezoelectricity, second-order permittivity.

Introduction

Background

The change in frequency of piezoelectric resonators caused by a dc electric field applied to their exciting electrodes is known as the polarizing effect. This phenomenon was studied experimentally in the past on a number of occasions with several types of piezo-

electric materials and for various modes of vibrations. As regards α -quartz, more thorough investigations have been reported for the longitudinal mode of thin rods and for the three thickness modes of thin plates.

Experimental Results : In the case of longitudinal vibrations of thin rectangular rods rotated about their thickness parallel to the crystallographic axis X, the change in series resonance frequency was found to be a linear function of the applied dc electric field with the field intensity as high as $4 \times 10^6 \text{ Vm}^{-1}$. Further, the change in frequency was found to be proportional to the magnitude of the fundamental frequency and to the order of the harmonic, and dependent on the crystallographic orientation of the cuts. A slight dependence of the frequency shift on temperature was also detected.^{1,2} Similar results were obtained later for the rods rotated about their widths parallel to the crystallographic axis Y and for doubly rotated rods with their thicknesses normal to the AT and BT planes.³

Qualitatively similar observations of the polarizing effect were also reported for the two fundamental thickness shear modes of a variety of doubly rotated planoconvex plates by Kusters⁴; and for the three fundamental thickness modes of thin plates rotated about one side of their square faces in the direction of the crystallographic axis Z.⁵ Kusters⁴ also recorded a strong temperature dependence of the effect and a quasi-exponential decay of the frequency change while the dc field was still applied.

In general, for both the rods and the plates, the typical magnitude of the change in frequency was found to be in the range of several parts per million at the dc electric field intensity of 10^6 Vm^{-1} .

Interpretation of the Observations : A first attempt at an interpretation of the polarizing effect was made for the longitudinal mode of thin rods. A simple formula, giving the frequency of vibrations as a function of the length, the density, and the elastic compliance in the length direction of the rod, was used. The changes in frequency were too large to be interpreted as a consequence of the change in length and density of the rods due to the converse piezoelectric effect. The introduction of a fifth rank electroelastic tensor related to the elastic compliances resulted in the determination of several of its components, in the reproduction of the experimental data with an excellent fit, and in remarkably accurate predictions for the magnitude of the polarizing effect for resonators which had not been hitherto used. These results were presented in a series of papers, of which Ref. 3 is the most comprehensive.

Several early papers dealing with the polarizing effect with the thickness vibrations of plates and essentially introducing the electroelastic tensor related to the elastic stiffnesses are represented by Ref. 6. A more thorough study of the causes of the polarizing effect with thickness vibrations of α -quartz plates can be found in Ref. 4. High internal stresses generated in the material of the plates by

the dc electric field are also considered in the paper along with the possible effect on the frequency of the third-order elastic constants. However, the corresponding frequency shifts are estimated to be smaller by an order of magnitude than those observed experimentally. All eight independent components of the electroelastic tensor related to the elastic stiffnesses are calculated from a frequency expression which includes only the effective elastic constants for the acoustic mode of interest. The possibility of the participation of other second-order effects is clearly recognized, as is the fact that in the given presentation these second-order effects would remain hidden in the calculated electroelastic tensor components.

A qualitatively improved interpretation of the polarizing effect dealing with the thickness vibrations of plates was presented in Ref. 5. It represented an improvement in the sense that it took into account the piezoelectric stiffening of the elastic constants, the coupling among the modes at the faces of the plates, and the changes in the amplitudes of the vibrations associated with the changes in the material constants in the dc electric field. The number of second-order effects considered was increased from one to three. On the other hand, the approach to the problem remained quasilinear in character. The polarizing effect was interpreted by means of linear changes in only those material constants that are present in the linear theory of vibrations. It was noted that this approach neglected other effects that could naturally appear in a nonlinear treatment. This interpretation was applied to data on the polarizing effect pertaining to all three thickness modes.⁵ Several electroelastic tensor components were determined. The attempt made to calculate the components of the tensors of second-order piezoelectricity and second-order permittivity resulted in failure. However, the electroelastic tensor components combined with the converse piezoelectric effect only reproduced the experimental data quite well.

Baumhauer and Tiersten published an interpretation of the polarizing effect based on the nonlinear equations of state.⁷ They concentrate on the thickness vibrations of a plate situated between parallel electrodes at a distance from the surfaces of the plate. The nonlinear approach and the implied conceptual reliability appear to be the advantages of this work. The resulting formulas however, cannot readily be applied to the existing experimental data.

Objectives and Method of the Present Work

Given the success of the simple work done on the polarizing effect with the longitudinal mode of thin rods on the one hand³, and the not-so-satisfactory results obtained for the thickness vibrations of plates on the other⁵, it was felt that the potential of the quasilinear interpretation of the polarizing effect had not been fully exploited. In the present paper the quasilinear interpretation⁵ is applied to a larger body of experimental data with the aim of

- 1) establishing that the polarizing effect can be used to study several second-order effects in α -quartz with an accuracy which makes the numerical results meaningful,
- 2) examining the size and the relative importance of the contributions of the second-order effects to the polarizing effect,
- 3) demonstrating that a numerically acceptable interpretation of the polarizing effect is possible even with the formal exclusion of all second-order

effects but electroelasticity, and

- 4) estimating the errors in the calculated electroelastic tensor components that occur when all effects contributing to the frequency changes, with the exception of the converse piezoelectric effect, are combined in the components of this tensor.

The work begins with a description of the method of measurement, of the data collected, and of the expression for the magnitude of the polarizing effect which will be used. The experimental data are then divided into two sets: one formed by the results pertaining to plates of orientation (xz1) ψ , and the other by those pertaining to plates of orientation (xzw) ϕ .

Using the first set of data and the adopted formula for the interpretation of the polarizing effect, several components of the tensors of electroelasticity, second-order piezoelectricity, and second-order permittivity are computed, and a comparison of their relative importance in the context of the polarizing effect is made. The calculations are then repeated without the latter two effects, and a new set of corresponding values for the electroelastic tensor components is found. The last calculations are also made independently for the second set of experimental data, and the values for the electroelastic tensor components obtained are compared with the values based on the first set. Also, using the second data set, a prediction is made about the first data set.

All the computed second-order constants are used to demonstrate their ability to reproduce the original experimental data.

The work is concluded with a brief discussion centering on the comparison of earlier measurements of the electroelastic tensor with the present results, and on the possible errors in their values.

Measurements and Their Results

Twenty-six resonators made of natural α -quartz were investigated. They were in the shape of thin square plates of side 25.000 mm and of thickness 0.600 mm. They were divided into two sets: one consisting of twelve plates of orientation (xz1) ψ with the cut angles ψ between 5° and 30° , the other consisting of fourteen plates of orientation (xzw) ϕ with the cut angles ϕ in the domain between -10° and -40° . The orientations are given according to the IRE definition.⁸ The resonators were plated with square electrodes, the side of the square being 18 mm. Beveling was used to avoid coupling with spurious modes. The plates were mounted by means of four pieces of thread forming a bipyramidal suspension.

Each resonator was continuously vibrated in a laboratory-built Heegner oscillator and functioned as an active part of the oscillator for the duration of the whole experiment. Prior to starting the experiment the oscillator was tuned to maximize the current through the resonator. With the oscillator tuned, the excitation level of the resonator was minimized to just sustain its vibration. The frequency of the resonator was understood to be the resonance frequency f_r in the sense of the IRE definition,⁹ the frequency being somewhat affected by the presence of the rest of the oscillator circuitry.

In order to expose a vibrating resonator to a dc electric field, a dc power supply was connected to the resonator electrodes. Two resistors separated the dc power supply from the oscillator to prevent short-circuiting of the high frequency of oscillations across

the dc power supply circuit parallel to the resonator. On the other hand, two blocking capacitors protected the Heegner oscillator from the dc potential across the resonator electrodes.

The frequency of the resonator oscillations and the frequency changes due to the applied dc electric field were measured by a Marconi counter/frequency meter, model 1417A.

The magnitude of the dc electric field in the resonators was taken as the ratio of the dc potential applied to the exciting electrodes to the nominal thickness of the resonators, the change in the thickness due to the converse piezoelectric effect being negligible.

The measurements were carried out in an air atmosphere and at a temperature of 20°C.

The change Δf in the natural resonance frequency f due to the applied dc electric field ΔE was measured for the three fundamental thickness modes of the resonators. The change in frequency Δf had been known to be dependent on the dc electric field intensity, on the crystallographic orientation of the resonators, and, for each resonator, on the mode of vibration.^{4,5} The linearity of the dependence of the frequency change Δf on the dc field ΔE was verified over the dc electric field range between $\pm 10^6$ Vm⁻¹. Because of this linear relationship, the magnitude of the polarizing effect could be conveniently expressed in terms of $(1/f) \cdot (\Delta f / \Delta E)$. For the twenty-six resonators investigated, these values were calculated as average of readings taken at $\pm 10^6$ Vm⁻¹. Figures 1 and 2 give the values of $(1/f) \cdot (\Delta f / \Delta E)$ for the two resonator sets plotted as functions of their respective cut angles ψ and ϕ .

The resonators were an unknown mixture of left- and right-hand quartz. Adjustments of the signs of the experimental data were made by means of the polarity of the charge which was measured when the plates were subject to a compressive stress in thickness. All the results in Figures 1 and 2 were then plotted in terms of the left-hand quartz.

Taking into account the existing random errors in the measurement of the natural resonance frequency f , in the frequency change Δf , the applied dc potential ΔV , and the tolerance in the resonator thickness t , the relative probable error in the values of $(1/f) \cdot (\Delta f / \Delta E)$ was found to vary between 2 and 3 percent. An additional error would be caused by the variations in the true orientation angle of the resonators with respect to their nominal orientation.

The quasi-exponential decay of the frequency change Δf taking place during observations while the dc electric field is applied must also be considered as a source of errors. Given the conditions under which the experiments described here were conducted, and using the available data on the Δf -decay for guidance,⁴ the error in the values of $(1/f) \cdot (\Delta f / \Delta E)$ caused by the drop in the frequency change Δf was estimated to be in the neighbourhood of 1 to 2 percent.

Another possible systematic error could arise from an inhomogeneity of the dc electric field. The accuracy of the readings could also be impaired by a possible presence of parasitic resonances.

Quasilinear Interpretation of the Polarizing Effect

General Formula

The expression used in this paper for the magnitude of the polarizing effect with the three thickness modes of thin plates was derived earlier.⁵ According to Eq. (40)⁵ one has

$$(1) \quad (1/f) \cdot (\Delta f / \Delta E) = {}^1\Lambda_{nrs} d_{nrs} + {}^2\Lambda_{ijkln} \delta_{ijkln} + {}^3\Lambda_{ijmn} q_{ijmn} + {}^4\Lambda_{rsn} t_{rsn},$$

where $i, j, k, l, m, n, r, s = 1, 2, 3$. The Einstein summation convention applies to these indices here and throughout the rest of the paper.

In Eq. (1), the quantities d_{nrs} are the piezoelectric strain constants. The quantities δ_{ijkln} , q_{ijmn} , and t_{nrs} are the components of the electro-elastic tensor, of second-order piezoelectricity, and of the tensor of second-order permittivity respectively. They are defined as follows

$$(2) \quad \delta_{ijkln} = \frac{\partial c_{ijkl}}{\partial E_n},$$

$$(3) \quad q_{ijmn} = \frac{\partial e_{mij}}{\partial E_n},$$

$$(4) \quad t_{nrs} = \frac{\partial \epsilon_{rs}}{\partial E_n}.$$

These partial derivatives of the elastic stiffnesses c_{ijkl} , of the piezoelectric stress constants e_{mij} , and of the dielectric permittivities ϵ_{rs} with respect to the components of the superposed dc electric field E_n are understood to be taken at zero dc electric field, at zero external stress and at constant entropy.

The coefficients ${}^1\Lambda_{nrs}$, ${}^2\Lambda_{ijkln}$, ${}^3\Lambda_{ijmn}$, and ${}^4\Lambda_{nrs}$ in Eq. (1) are functions of the elastic stiffnesses, of the piezoelectric stress constants and of the dielectric permittivities of the crystalline material. Furthermore, they depend on the crystallographic orientation of the plate, on its thickness, on the electrode separation, on the mode of vibration, on the order of harmonic, and on the direction of the amplitudes of the vibrations.⁵ They do not depend explicitly on the excitation level of the plate vibrations or on the temperature.

The expression (1) for the magnitude of the polarizing effect is based on Eqs. (2)-(5).⁵ These equations represent the conditions which obtain in the linear approximation for the eigenfrequencies and the amplitudes of the thickness modes of piezoelectric plates of infinite lateral dimensions placed between infinite plane electrodes. In order to obtain Eq. (1), these conditions were differentiated with respect to the superposed dc electric field E treated as a parameter. The differentiated conditions were then solved for $(1/f) \cdot (\partial f / \partial E)$ at $E = 0$. Subsequently, the obtained solution was put equal to the measured quantity $(1/f) \cdot (\Delta f / \Delta E)$. This was possible because the measurements were carried out in the linear domain of the polarizing effect. Accordingly, the tensor quantities defined by Eqs. (2)-(4) appeared in Eq. (1) as a result of this process rather than as the members of

the nonlinear equations of state.

The symbols $1\Lambda_{nrs}$, $2\Lambda_{ijkln}$, $3\Lambda_{ijmn}$, and $4\Lambda_{nrs}$ found in Eq. (1) were not used in Ref. 5. Together with some other minor notational changes they are introduced here to simplify further treatment of Eq. (1) and its application to α -quartz. In relation to the notation used before⁵, the following definitions apply

$$(5) \quad 1\Lambda_{nrs} = \frac{1}{2L} \left[5L_{ik} \delta_{ir} \delta_{ks} + 3L_{ik} \frac{F_i F_k}{\epsilon} \left(\frac{\epsilon d}{b} - \frac{\epsilon t}{b} \delta_{(d-t)} \right) \alpha_r \alpha_s \right] \alpha_n,$$

$$(6) \quad 2\Lambda_{ijkln} = \frac{1}{2L} \left[L_{ik} - 2L_{ik} - 4L_{ik} \right] \alpha_j \alpha_l \alpha_n,$$

$$(7) \quad 3\Lambda_{ijmn} = \frac{1}{L} \left[L_{ik} - 2L_{ik} + 3L_{ik} - 4L_{ik} \right] \frac{F_k}{\epsilon} \alpha_j \alpha_m \alpha_n,$$

$$(8) \quad 4\Lambda_{nrs} = \frac{1}{2L} \left[-L_{ik} + 2L_{ik} - 3L_{ik} \left(1 + \frac{\epsilon(d-t)}{b} \right) + 4L_{ik} \right] \frac{F_i F_k}{\epsilon^2} \alpha_n \alpha_r \alpha_s,$$

and further

$$(9) \quad L = \sum_{z=1}^3 E_{iz} A_k(z) \Gamma_{ik} \left(B(z) \cot B(z) - \frac{(B(z))^2}{\sin^2 B(z)} \right)$$

$$(10) \quad 1L_{ik} = \sum_{z=1}^3 E_{rz} A_s(z) \Gamma_{rs} \left(B(z) \cot(z) - \frac{(B(z))^2}{\sin^2 B(z)} \right) \frac{1}{\Gamma(z)} D_{ik}(z),$$

$$(11) \quad 2L_{ik} = \sum_{z=1}^3 \left(E_{iz} A_k(z) + E_{kz} A_i(z) \right) B(z) \cot B(z)$$

$$(12) \quad 3L_{ik} = \frac{\epsilon_0 t}{b} \sum_{z=1}^3 \left(E_{iz} A_k(z) + E_{kz} A_i(z) \right),$$

$$(13) \quad 4L_{ik} = \sum_{z=1}^3 E_{rz} \left(\Gamma_{rs} B(z) \cot B(z) - \frac{\epsilon_0 t}{b} \frac{F_r F_s}{\epsilon} \right) \left(M_{sik}(z) + M_{skl}(z) \right),$$

$$(14) \quad 5L_{ik} = \left(-\alpha_i \alpha_k + \beta_i \beta_k + \gamma_i \gamma_k \right) L$$

As regards all other symbols used in Eqs. (5)-(14), their earlier definitions⁵ remain in effect throughout this paper.

All material constants appearing in Eq. (1) as well as the quantities defined by Eqs. (2)-(4) are related to the basic frame of reference defined in piezoelectric crystals.⁸

According to Eqs. (5)-(8), each of the coefficients $1\Lambda_{nrs}$, $2\Lambda_{ijkln}$, $3\Lambda_{ijmn}$, and $4\Lambda_{nrs}$ is an algebraic sum of two or more terms, each containing one factor of the type PL_{ik} , $p = 1, 2, \dots, 5$. The factors PL_{ik} are characteristic of the origin of these terms. They relate each of the terms to a specific part of the condition for frequency of the thickness modes of plates in Eq. (2)⁶ and to the quantitative change this part is assumed to undergo as a result of the applied dc electric field. In this sense the terms containing $1L_{ik}$, $2L_{ik}$, $3L_{ik}$, $4L_{ik}$, and $5L_{ik}$ correspond, in the same order, to the effective elastic moduli of the plate $\Gamma(z)$ in Eq. (10)⁵, to the quantities Γ_{ik} in Eq. (7)⁶, to the terms $\epsilon_0 t F_i F_k / \epsilon b$ representing the coupling among the modes at the traction free surfaces of the plate in Eq. (7)⁶, to the direction cosines $A_i(z)$ of the amplitudes of vibrations in Eq. (7)⁶, and to the density ρ and the thickness t of the plate in $B(z)$ in Eq. (10).⁵

The large number of terms in Eq. (1) can be substantially reduced by means of the traditional matrix index notation.¹⁰ By replacing the pairs of Latin indices ij , kl , and rs by a single Greek index λ or μ , $\lambda, \mu = 1, 2, \dots, 6$, one obtains

$$(15a) \quad d_{n\lambda} = d_{nrs}, \quad \text{for } r = s,$$

$$(15b) \quad d_{n\lambda} = 2 d_{nrs} = 2 d_{nsr}, \quad \text{for } r \neq s,$$

$$(16) \quad \delta_{\lambda\mu n} = \delta_{ijkln} = \delta_{ijlkn} = \delta_{jikln} = \delta_{jilkn},$$

$$(17) \quad q_{\lambda\mu} = q_{ijmn} = q_{ijnm} = q_{jimn} = q_{jinm},$$

where the substitution is made using the standard rules.¹⁰

Application of the Kleinman conditions¹¹ leads to a contraction of the second-order permittivity tensor. A single Greek index Δ can be used to replace each triad of the Latin indices nrs consisting of the same index values irrespective of their order according to the following rule

$$(18) \quad \Delta = n . r . s$$

so that the new components t_Δ can be introduced as shown

$$(19) \quad t_\Delta = t_{nrs} = t_{nsr} = t_{snr} = t_{srn} = t_{rsn} = t_{rns}$$

The set of 10 index values Δ generated by Eq. (18) is "discontinuous", which is somewhat unusual. On the other hand the Greek indices Δ can be easily related to the corresponding original triad of the Latin tensor indices. In Table I the correspondence between the indices Δ and the triads nrs is shown in detail.

To take advantage of the contraction of tensors described in Eqs. (15a)-(17) and (19), the coefficients $1\Lambda_{nrs}$, $2\Lambda_{ijkln}$, $3\Lambda_{ijmn}$, and $4\Lambda_{nrs}$ must be adjusted accordingly. Using the same rules of assignment of the Greek indices, the following definitions are convenient

$$(20) \quad 1\Lambda_{n\lambda} = 1\Lambda_{nrs},$$

$$(21a) \quad 2\Lambda_{\lambda\mu n} = 2\Lambda_{ijkln}, \quad \text{for } i = j, k = l,$$

$$(21b) \quad {}^2\lambda_{\lambda\mu n} = {}^2\lambda_{ijkln} + {}^2\lambda_{ijlkn}, \quad \text{for } i = j, k \neq l,$$

$$(21c) \quad {}^2\lambda_{\lambda\mu n} = {}^2\lambda_{ijkln} + {}^2\lambda_{jikln}, \quad \text{for } i \neq j, k = l,$$

$$(21d) \quad {}^2\lambda_{\lambda\mu n} = {}^2\lambda_{ijkln} + {}^2\lambda_{ijlkn} + {}^2\lambda_{jikln} + {}^2\lambda_{jilkkn}, \\ \text{for } i \neq j, k \neq l,$$

$$(22a) \quad {}^3\lambda_{\lambda\mu} = {}^3\lambda_{ijmn}, \quad \text{for } i = j = m = n,$$

$$(22b) \quad {}^3\lambda_{\lambda\mu} = {}^3\lambda_{ijmn} + {}^3\lambda_{ijnm}, \quad \text{for } i = j, m \neq n,$$

$$(22c) \quad {}^3\lambda_{\lambda\mu} = {}^3\lambda_{ijmn} + {}^3\lambda_{jimn}, \quad \text{for } i \neq j, m = n,$$

$$(22d) \quad {}^3\lambda_{\lambda\mu} = {}^3\lambda_{ijmn} + {}^3\lambda_{ijnm} + {}^3\lambda_{jimn} + {}^3\lambda_{jinm}, \\ \text{for } i \neq j, m \neq n,$$

$$(23a) \quad {}^4\lambda_{\Delta} = {}^4\lambda_{nrs}, \quad \text{for } n = r = s,$$

$$(23b) \quad {}^4\lambda_{\Delta} = {}^4\lambda_{nrs} + {}^4\lambda_{rsn} + {}^4\lambda_{snr}, \quad \text{for } n = r \neq s \text{ or} \\ \text{for } r = s \neq n \text{ or} \\ \text{for } s = n \neq r,$$

$$(23c) \quad {}^4\lambda_{\Delta} = {}^4\lambda_{nrs} + {}^4\lambda_{nsr} + {}^4\lambda_{rsn} + {}^4\lambda_{rns} + {}^4\lambda_{snr} + {}^4\lambda_{srn}, \\ \text{for } n \neq r \neq s.$$

Using the new index notation, Eq. (1) takes on the following form

$$(24) \quad (1/f) \cdot (\Delta f / \Delta E) = {}^1\lambda_{n\lambda} d_{n\lambda} + {}^2\lambda_{\lambda\mu n} \delta_{\lambda\mu n} + {}^3\lambda_{\lambda\mu} q_{\lambda\mu} + {}^4\lambda_{\Delta} t_{\Delta}$$

where $n = 1, 2, 3$, $\lambda, \mu = 1, 2, \dots, 6$, and $\Delta = 1, 2, 3, 4, 6, 8, 9, 12, 18, 27$. The Einstein summation rule remains in effect for all index values listed.

The new index notation results in a substantial reduction in the number of terms on the right hand side of Eq. (1): from the original number of 378 in Eq. (1) to only 172 in Eq. (24). This number can be further reduced to only 127 terms by means of the relationship

$$(25) \quad \delta_{\lambda\mu n} = \delta_{\mu\lambda n}, \quad \lambda, \mu = 1, 2, \dots, 6,$$

that follows from the thermodynamic considerations.

Application to α -quartz

In the above form Eqs. (24) and (25) would apply to the crystalline material of any symmetry. In the case of α -quartz Eq. (24) is further simplified, as the number of the independent tensor components $d_{n\lambda}$, $\delta_{\lambda\mu n}$, $q_{\lambda\mu}$, and t_{Δ} is reduced due to the crystal symmetry of α -quartz. When Eqs. (24) and (25) are combined, the following result is obtained

$$(26) \quad (1/f) \cdot (\Delta f / \Delta E) = {}^1\lambda_{11}^Q d_{11} + {}^1\lambda_{14}^Q d_{14} \\ + {}^2\lambda_{111}^Q \delta_{111} + {}^2\lambda_{131}^Q \delta_{131} + {}^2\lambda_{141}^Q \delta_{141} \\ + {}^2\lambda_{221}^Q \delta_{221} + {}^2\lambda_{241}^Q \delta_{241} + {}^2\lambda_{341}^Q \delta_{341} \\ + {}^2\lambda_{441}^Q \delta_{441} + {}^2\lambda_{153}^Q \delta_{153} + {}^3\lambda_{11}^Q q_{11} \\ + {}^3\lambda_{12}^Q q_{12} + {}^3\lambda_{13}^Q q_{13} + {}^3\lambda_{14}^Q q_{14} + {}^3\lambda_{31}^Q q_{31} \\ + {}^3\lambda_{33}^Q q_{33} + {}^3\lambda_{41}^Q q_{41} + {}^3\lambda_{44}^Q q_{44} + {}^4\lambda_{11}^Q t_{11},$$

where

$$(27a) \quad {}^1\lambda_{11}^Q = {}^1\lambda_{11} - {}^1\lambda_{12} - 2{}^1\lambda_{26},$$

$$(27b) \quad {}^1\lambda_{14}^Q = {}^1\lambda_{14} - {}^1\lambda_{25},$$

$$(28a) \quad {}^2\lambda_{111}^Q = {}^2\lambda_{111} - \frac{1}{2}({}^2\lambda_{121} + {}^2\lambda_{211}) - \frac{1}{2}{}^2\lambda_{661} \\ + \frac{1}{4}({}^2\lambda_{162} + {}^2\lambda_{612}) - \frac{3}{4}({}^2\lambda_{262} + {}^2\lambda_{622}),$$

$$(28b) \quad {}^2\lambda_{131}^Q = {}^2\lambda_{131} + {}^2\lambda_{311} - {}^2\lambda_{231} - {}^2\lambda_{321} - {}^2\lambda_{361} - {}^2\lambda_{631},$$

$$(28c) \quad {}^2\lambda_{141}^Q = {}^2\lambda_{141} + {}^2\lambda_{411} - \frac{1}{2}({}^2\lambda_{561} + {}^2\lambda_{651}) - ({}^2\lambda_{252} + {}^2\lambda_{522}) \\ + \frac{1}{2}({}^2\lambda_{462} + {}^2\lambda_{642}),$$

$$(28d) \quad {}^2\lambda_{221}^Q = {}^2\lambda_{221} - \frac{1}{2}({}^2\lambda_{121} + {}^2\lambda_{211}) - \frac{1}{2}{}^2\lambda_{661} \\ + \frac{3}{4}({}^2\lambda_{162} + {}^2\lambda_{612}) - \frac{1}{4}({}^2\lambda_{262} + {}^2\lambda_{622}),$$

$$(28e) \quad {}^2\lambda_{241}^Q = {}^2\lambda_{241} + {}^2\lambda_{421} + \frac{1}{2}({}^2\lambda_{561} + {}^2\lambda_{651}) - {}^2\lambda_{152} - {}^2\lambda_{512} \\ - \frac{1}{2}({}^2\lambda_{462} + {}^2\lambda_{642}),$$

$$(28f) \quad {}^2\lambda_{341}^Q = {}^2\lambda_{341} + {}^2\lambda_{431} - {}^2\lambda_{352} - {}^2\lambda_{532},$$

$$(28g) \quad {}^2\lambda_{441}^Q = {}^2\lambda_{441} - {}^2\lambda_{551} + {}^2\lambda_{452} + {}^2\lambda_{542},$$

$$(28h) \quad {}^2\lambda_{153}^Q = {}^2\lambda_{153} + {}^2\lambda_{513} - {}^2\lambda_{253} - {}^2\lambda_{523} - {}^2\lambda_{463} - {}^2\lambda_{643},$$

$$(29a) \quad {}^3\lambda_{11}^Q = {}^3\lambda_{11} + {}^3\lambda_{22} + \frac{1}{2}{}^3\lambda_{66},$$

$$(29b) \quad {}^3\lambda_{12}^Q = {}^3\lambda_{12} + {}^3\lambda_{21} - \frac{1}{2}{}^3\lambda_{66},$$

$$(29c) \quad {}^3\lambda_{13}^Q = {}^3\lambda_{13} + {}^3\lambda_{23},$$

$$(29d) \quad {}^3\Lambda_{14}^Q = {}^3\Lambda_{14} - {}^3\Lambda_{24} + {}^3\Lambda_{65},$$

$$(29e) \quad {}^3\Lambda_{31}^Q = {}^3\Lambda_{31} + {}^3\Lambda_{32},$$

$$(29f) \quad {}^3\Lambda_{41}^Q = {}^3\Lambda_{41} - {}^3\Lambda_{42} + {}^3\Lambda_{56},$$

$$(29g) \quad {}^3\Lambda_{33}^Q = {}^3\Lambda_{33},$$

$$(29h) \quad {}^3\Lambda_{44}^Q = {}^3\Lambda_{44} + {}^3\Lambda_{55},$$

$$(30) \quad {}^4\Lambda_1^Q = {}^4\Lambda_1 - {}^4\Lambda_4.$$

In Eq. (26) we find the magnitude of the polarizing effect with the thickness vibrations of α -quartz plates as a linear function of the independent components of the participating tensors. These are: 2 components $d_{n\lambda}$ of the tensor of the converse piezoelectric effect, 8 components $\delta_{\lambda\mu n}$ of the tensor of the electroelastic effect, 8 components $q_{\lambda\mu}$ of the tensor of the second-order piezoelectricity, and 1 component $t_\lambda = t_1$ of the tensor of the second-order permittivity. In order to obtain Eqs. (26) and (27a)-(30), the matrix diagrams of these four tensors are needed in full detail.^{4,10,12,13} (The matrix diagram of the photoelastic coefficients p_{ijmn} must be referred to¹⁰ in order to obtain the structure of the second-order piezoelectricity tensor).

The values of the coefficients of the types ${}^1\Lambda_{n\lambda}^Q$, ${}^2\Lambda_{\lambda\mu n}^Q$, ${}^3\Lambda_{\lambda\mu}^Q$, and of the coefficient ${}^4\Lambda_1^Q$ in Eq. (26) were calculated for a number of plate orientations, and for different plate thicknesses, electrode separations and orders of the harmonic of the three thickness modes. The results obtained indicate that the values of these coefficients depend primarily on the plate orientation. The changes in the remaining factors affect the coefficients in Eq. (26) only to the extent of several percent of their values. The maximum orders of magnitude of the coefficients ${}^1\Lambda_{n\lambda}^Q$, ${}^2\Lambda_{\lambda\mu n}^Q$, ${}^3\Lambda_{\lambda\mu}^Q$, and ${}^4\Lambda_1^Q$ were found to be 10^{-1} , 10^{-12} N⁻¹m², 10^{-3} VN⁻¹m, and 10^7 F⁻¹m, respectively.

The values of the elastic stiffnesses $c_{\lambda\mu}$, of the piezoelectric stress constants $e_{n\lambda}$, and of the dielectric permittivities ϵ_{mn} needed to calculate the above coefficients were taken from Ref. 14. The value $\rho = 2.649 \times 10^3$ kg.m⁻³ was used for the density of α -quartz.

Further analysis of the values of the coefficients ${}^1\Lambda_{n\lambda}^Q$, ${}^2\Lambda_{\lambda\mu n}^Q$, ${}^3\Lambda_{\lambda\mu}^Q$, and ${}^4\Lambda_1^Q$ in Eq. (26) was undertaken using Eqs. (5)-(8). In Eq. (5) the term containing the factor ${}^5L_{ik}$ was found to be always larger, by two orders of magnitude, than the term containing the factor ${}^3L_{ik}$. In Eqs. (6)-(8) the factors ${}^1L_{ik}$ and ${}^3L_{ik}$ are of the same order of magnitude and about 100 times larger than the factors ${}^2L_{ik}$ and ${}^4L_{ik}$. As far as the factor ${}^3L_{ik}$ is concerned, this implies that if the interpretation of the polarizing effect includes the second-order piezoelectricity and the second-order permittivity, the coupling among the thickness modes, although it complicates the calculations considerably, should not be neglected. At the same time the dominant positions of ${}^1L_{ik}$ in Eq. (6) and of ${}^5L_{ik}$ in Eq. (5) explain the agreement between our present results and those obtained by Kusters⁴ which will be shown later.

Two sets of plates are investigated experimentally in the present paper. For the first set of plates of orientation (xzl) ψ , the following coefficients in Eq. (26) are found to be equal to zero for all cut angles ψ

$$(31) \quad {}^1\Lambda_{14}^Q = 0,$$

$$(32) \quad {}^2\Lambda_{131}^Q = {}^2\Lambda_{341}^Q = {}^2\Lambda_{153}^Q = 0,$$

$$(33) \quad {}^3\Lambda_{12}^Q = {}^3\Lambda_{13}^Q = {}^3\Lambda_{14}^Q = {}^3\Lambda_{31}^Q = {}^3\Lambda_{33}^Q = {}^3\Lambda_{44}^Q = 0.$$

As regards the second set of plates of orientation (xzw) ψ , the following holds true in general

$$(34) \quad {}^1\Lambda_{14}^Q = 0,$$

$$(35) \quad {}^3\Lambda_{12}^Q = 0.$$

In addition to the above equations, and for each set of the plates, several nonzero coefficients in Eq. (26) are found to be linearly dependent on other nonzero coefficients, eg.

$$(36) \quad {}^2\Lambda_{441}^Q = -1.00 {}^2\Lambda_{141}^Q + 2.00 {}^2\Lambda_{221}^Q$$

which holds for all plates of orientation (xzl) ψ . Eqs. (31)-(35) and relationships of the type shown in Eq. (36) simplify Eq. (26) considerably. Unfortunately, they also reduce the amount of information about the second-order effect that can be obtained from the experimental data available.

Second-Order Effects

Electroelastic Effect, Second-Order Piezoelectricity and Second-Order Permittivity

In order to demonstrate the possibility of determining the components of the tensors of the electroelasticity, of the second-order piezoelectricity and of the second-order permittivity, Eq. (26) was applied to the 32 experimental data available for the twelve plates of orientation (xzl) ψ . Eqs. (31)-(33) were used and Eq. (26) was rewritten as follows

$$(37) \quad \frac{1}{f} \frac{\Delta f}{\Delta E} - {}^1\Lambda_{11}^Q d_{11} = {}^2\Lambda_{111}^Q \delta_{111} + {}^2\Lambda_{141}^Q \delta_{141} + {}^2\Lambda_{221}^Q \delta_{221} + {}^2\Lambda_{241}^Q \delta_{241} + {}^2\Lambda_{441}^Q \delta_{441} + {}^3\Lambda_{11}^Q q_{11} + {}^3\Lambda_{41}^Q q_{41} + {}^4\Lambda_1^Q t_1.$$

The substitution of the available experimental values for the magnitude of the polarizing effect ($1/f$), $(\Delta f/\Delta E)$ and of the known¹⁴ piezoelectric strain constant d_{11} resulted in a system of 32 simultaneous equations for the eight second-order tensor components δ_{111} , δ_{141} , δ_{221} , δ_{241} , δ_{441} , q_{11} , q_{41} , and t_1 .

Eq. (36) and another two linear relationships were found to hold among the coefficients ${}^2\Lambda_{111}^Q$,

$2\Lambda_{141}^Q, \dots$, and $4\Lambda_1^Q$ in Eq. (37) for all the cut angles ψ and the acoustic modes studied experimentally. After the inclusion of these relationships into Eq. (37), each equation of our system took on the following form

$$(38) \quad (1/f) \cdot (\Delta f / \Delta E) - \Lambda_{11}^Q d_{11} = \\ 2\Lambda_{111}^Q (\delta_{111} + 1.63 \times 10^9 q_{11}) \\ + 2\Lambda_{141}^Q (\delta_{141} - \delta_{241} - 1.00 \delta_{441}) \\ + 2\Lambda_{221}^Q (\delta_{221} + 2.00 \delta_{441} - 1.62 \times 10^9 q_{11}) \\ + 3\Lambda_{41}^Q q_{41} + 4\Lambda_1^Q t_1.$$

The dimensions of the numerical coefficients 1.63×10^9 and 1.62×10^9 are Vm^{-1} .

The method of least squares was used to find the suitable values for the five second-order tensor components and their combinations appearing on the right hand side of Eq. (38). The results obtained are presented in Table II together with their mean probable errors.

The calculated components and component combinations of the tensors of electroelasticity, second-order piezoelectricity and second-order permittivity listed in Table II were substituted in Eq. (38) in order to demonstrate their ability to reproduce the thirty two original experimental values of $(1/f) \cdot (\Delta f / \Delta E)$. While most of the values themselves are of the order of $10^{-11} \text{V}^{-1} \text{m}$ or $10^{-12} \text{V}^{-1} \text{m}$, the average absolute difference was only 1.5×10^{-13} , the maximum difference being $6.2 \times 10^{-13} \text{V}^{-1} \text{m}$. In terms of percentages, the average deviation of the calculated values of $(1/f) \cdot (\Delta f / \Delta E)$ was 2.8% of their experimental counterparts. Given that the experimental errors in $(1/f) \cdot (\Delta f / \Delta E)$ were estimated to be about 4 percent, the agreement is good.

Eq. (38) can be applied to the fundamental frequency of the three thickness modes of the plates in the whole domain of orientations $(xzl)\psi$. Using the quantities in Table II, calculations of the magnitude of the polarizing effect $(1/f) \cdot (\Delta f / \Delta E)$ as a function of the cut angle ψ were made for $0^\circ \leq \psi \leq 30^\circ$ at 5° intervals. The three curves obtained predicting the dependence of the polarizing effect on the cut angle ψ for the three acoustic modes are shown in Fig. 1. For the angles ψ , $30^\circ < \psi \leq 60^\circ$, the graphs are odd functions of ψ for $\psi = 30^\circ$. The experimental values for $(1/f) \cdot (\Delta f / \Delta E)$ which are also plotted in Fig. 1 give an overall picture of the relationship between the experimental data and the fitted curves.

Using the computed value of q_{41} for guidance in estimating the magnitude of the contribution of the second-order piezoelectricity, and considering it unlikely that the order of magnitude of the component q_{11} be greater than 10^{-9}NV^{-2} (Table II) one can conclude that the most important contributor to the polarizing effect $(1/f) \cdot (\Delta f / \Delta E)$ is the electroelastic tensor. When applied to a number of cut angles ψ , Eq. (38) showed that the contribution of the electroelastic effect was always of the order of the polarizing effect itself. With the polarizing effect exceeding $10^{-11} \text{V}^{-1} \text{m}$, the converse piezoelectric effect accounts only for multiples of $10^{-12} \text{V}^{-1} \text{m}$. The order of magnitude

of the contributions of the second-order permittivity is not greater than $10^{-13} \text{V}^{-1} \text{m}$. As regards the second-order piezoelectricity, its typical contribution is of the same order of magnitude. However, should the order of magnitude of q_{11} be 10^{-9}NV^{-2} , its contribution in some cases would be a comparable to that of the converse piezoelectric effect.

As regards the computed second-order tensor quantities containing the components of the electroelastic tensor $\delta_{\lambda\mu\nu}$, the probable errors stated in Table II are each smaller than 4%, the average being 3%. Since the experimental errors in the values of $(1/f) \cdot (\Delta f / \Delta E)$ are approximately the same, the 3% errors are consistent with the position of the electroelastic tensor in Eq. (38). The percentage errors are larger for the quantities q_{41} and t_1 . This result is understandable since q_{41} and t_1 play a qualitatively unimportant role in Eq. (38).

Cumulative Electroelastic Tensor

The dominant position of the electroelastic tensor implied in the previous section was confirmed when it was found that a successful numerical interpretation of the available experimental data was possible while formally neglecting the second-order piezoelectricity and the second-order permittivity.

The plates of orientation $(xzl)\psi$: The steps that had led from Eq. (26) to Eq. (38) were followed and the terms $3\Lambda_{111}^Q q_{11}$, $3\Lambda_{41}^Q q_{41}$, and $4\Lambda_1^Q t_1$ in Eq. (37) were disregarded. A new system of 32 simultaneous equations was obtained, relating the magnitude of the polarizing effect $(1/f) \cdot (\Delta f / \Delta E)$ to the electroelastic and the converse piezoelectric effects only. These equations were of the following type

$$(39) \quad (1/f) \cdot (\Delta f / \Delta E) - \Lambda_{11}^Q d_{11} = 2\Lambda_{111}^Q \delta_{111} \\ + 2\Lambda_{141}^Q (\delta_{141} - \delta_{241} - 1.00 \delta_{441}) \\ + 2\Lambda_{221}^Q (\delta_{221} + 2.00 \delta_{441})$$

The method of least squares was used again to compute the suitable values of the electroelastic tensor component δ_{111} and of the component combinations $\delta_{141} - \delta_{241} - \delta_{441}$ and $\delta_{221} + 2\delta_{441}$. The results of this calculation are shown in Table III in the column marked $(xzl)\psi$. The probable errors in the computed quantities are in the neighbourhood of 3%.

The quality of reproduction of the experimental data by means of the calculated tensor components is shown in Fig. 1. The fitted curves giving the dependence of the magnitude of the polarizing effect on the cut angle ψ for the three thickness modes are calculated using Eq. (39). The average absolute difference between the calculated and the measured values is $2.4 \times 10^{-13} \text{V}^{-1} \text{m}$, the maximum difference being $6.8 \times 10^{-13} \text{V}^{-1} \text{m}$. The agreement between the calculated and the experimental values is not substantially different from the one obtained when all second-order effects appearing in the quasilinear approximation of the polarizing effect are taken into account. At the same time it should be noted that the whole set of 32 experimental data is now being reproduced using only three calculated constants.

In connection with the above results it is relevant to mention that attempts have been made to use

Eq. (37) and to interpret the data for the magnitude of the polarizing effect while neglecting either completely or in part the terms ${}^2\lambda_{\mu\nu}^Q \delta_{\lambda\mu\nu}$. All these attempts failed completely. The least square estimates of the second-order tensor components retained in the final form of Eq. (37) were always found to be associated with large probable errors, and the original experimental data for the magnitude of the polarizing effect $(1/f) \cdot (\Delta f/\Delta E)$ were never reproduced satisfactorily.

While the three numerical values in Table III reproduce the experimental data well, they are not recommended as the best values for the electroelastic tensor components δ_{111} , $\delta_{141} - \delta_{241} - \delta_{441}$, and $\delta_{221} + 2\delta_{441}$. In Eq. (39), these components take over, at least in part, the function of the quantities q_{11} , q_{41} , and t_1 which are present in Eqs. (37) and (38) but are formally rejected in Eq. (39). As a result, it is possible that two or three second-order effects are combined in the values given in Table III. This is why we call the quantities calculated in this section the components of the cumulative electroelastic tensor.

The plates of orientation (xzw): One component and five component combinations of the cumulative electroelastic tensor were determined from the 29 data (Fig. 2) pertaining to the polarizing effect with the fourteen plates of orientation (xzw). The experimental data used in the preceding section were not involved. The values of the calculated tensor components together with their probable errors are given in Table IV in the column marked (xzw). They were calculated by the method of least squares from a system of 29 equations of the following type

$$\begin{aligned} (40) \quad (1/f) \cdot (\Delta f/\Delta E) - {}^1\lambda_{11}^Q \delta_{11} = & \\ & {}^2\lambda_{111}^Q \delta_{111} + {}^2\lambda_{131}^Q (\delta_{131} - 0.42 \delta_{221} + 0.25 \delta_{341}) \\ & {}^2\lambda_{141}^Q (\delta_{141} - 0.26 \delta_{221} - 0.67 \delta_{341}) \\ & + {}^2\lambda_{241}^Q (\delta_{241} - 0.77 \delta_{221} - 0.67 \delta_{341}) \\ & + {}^2\lambda_{441}^Q (\delta_{441} + 0.50 \delta_{221}) \\ & + {}^2\lambda_{153}^Q (\delta_{153} + 0.60 \delta_{221} - 0.51 \delta_{341}). \end{aligned}$$

Eq. (40) obtains from Eq. (26) when Eq. (26) is applied to any plate of orientation (xzw). Eqs. (34) and (35) are used, and all nonzero terms of the type ${}^1\lambda_{\mu\nu}^Q$ and the term ${}^4\lambda_{11}^Q t_1$ are disregarded. Two coefficients, ${}^2\lambda_{221}^Q$ and ${}^2\lambda_{341}^Q$, were found to be linearly dependent on other coefficients ${}^2\lambda_{\mu\nu}^Q$ in the whole domain of cut angles ϕ . These linear relationships were also included in Eq. (26) in order to yield Eq.

The ability of the obtained results to reproduce the experimental data is shown in Fig. 2. The curves giving the magnitude of the polarizing effect for all three thickness modes are shown in Fig. 3. The average absolute difference between the calculated and the measured values is 3.0 $\times 10^{-13}$ V $^{-1}$ m, the maximum difference is

The results from the column marked (xzw) in Table IV can be compared with the values for the cumulative electroelastic tensor components in Table III independently obtained before. A direct comparison is possible for the component δ_{111} ; the values for $\delta_{141} - \delta_{241} - \delta_{441}$ and for $\delta_{221} + 2\delta_{441}$ can be obtained from Table IV by combination of the available results. The comparison is made in Table III.

Moreover, the twenty-nine data on the polarizing effect with the plates of orientation (xzw) and the cumulative electroelastic tensor components calculated from them were used to predict directly the magnitude of the polarizing effect with the plates of orientation (xzl). The calculated curves and the experimental values obtained independently are shown in Fig. 4. For the thirty two experimental data at our disposal the maximum absolute difference between the predicted and the measured values is 1.08×10^{-12} V $^{-1}$ m; the average difference is 3.6×10^{-13} V $^{-1}$ m, the median only 3.0×10^{-13} V $^{-1}$ m.

All calculations in this section of the paper were made for the left-hand modification of α -quartz. Consequently, the values of the second-order tensor components in Tables II-IV are also related to the left quartz. If similar constants are required for the right-hand modification, one should change the sign of all constants $\delta_{\lambda\mu\nu}$ and of the constant t_1 . However, the signs of the constants $q_{\lambda\mu}$ should not be changed.

Conclusion

In the present paper the observed changes in frequency are associated with the changes which are induced by the dc electric field and assumed to take place in the quantities appearing in the linear theory of the thickness vibrations of α -quartz plates. The polarizing effect is thus interpreted as a consequence of the changes in the dimensions of the plate and the airgap due to the converse piezoelectric effect and the changes in the values of the elastic, piezoelectric, and dielectric constants. The changes in the material constants are described by means of a fifth rank tensor of electroelasticity, a fourth rank tensor of second-order piezoelectricity, and a third rank tensor of second-order permittivity. In the resulting Eq. (26), the magnitude of the polarizing effect also depends on the orientation of the plate, its thickness, the electrode separation, the acoustic mode, and the order of the harmonic.

In conjunction with the experimental values for the magnitude of the polarizing effect, Eq. (26) led to the determination of several components and component combinations of the tensors of the second-order effects. Insofar as one can conclude from the small number of values obtained, namely five, the order of magnitude of the electroelastic tensor components $\delta_{\lambda\mu\nu}$ is 10^0 NV $^{-1}$ m $^{-1}$, and that of the second-order piezoelectricity tensor components $q_{\lambda\mu}$ is 10^{-10} NV $^{-2}$. The only independent component t_1 of the second-order permittivity was found to be of the order of 10^{-20} FV $^{-1}$. The ability of these quantities to reproduce the original experimental data is better than 3% on the average.

The electroelastic effect appears to be primarily responsible for the frequency changes observed, namely for the larger values of the polarizing effect of the order of 10^{-11} Vm $^{-1}$. The converse piezoelectric effects ranks second, with contributions of the order of 10^{-12} V $^{-1}$ m. In general, the second-order piezoelectricity and the second-order permittivity seem to be of less importance. This view is supported by

several results obtained when the latter two effects were formally neglected, namely the numerically successful interpretation of the polarizing effect by means of the cumulative electroelastic tensor, the agreement between the values of its components when calculated from independent experimental data, and the accuracy of the predictions about the magnitude of the polarizing effect. The same view is also supported by a series of unsuccessful attempts to obtain similar results when the conditions were reversed i.e. when the tensors of the second-order piezoelectricity and of the second-order permittivity were retained while the electroelastic tensor was formally excluded from the calculations.

The electroelastic tensor was calculated on a few previous occasions using a simpler or somewhat different approach. The complete set of eight electroelastic tensor components presented by Kusters⁴ comes from an independent source. Considering the method of their calculation, they correspond very closely to the cumulative electroelastic tensor components in the present paper. Using their actual numerical values¹⁵, a comparison with our results is made in Table IV. If we realize that except for δ_{111} the order of magnitude of the components $\delta_{\lambda\mu\nu}$ in Ref. 4 is $10 \text{ NV}^{-1}\text{m}^{-1}$ to $10^2 \text{ NV}^{-1}\text{m}^{-1}$, the degree of agreement is surprisingly good.

The magnitude of the probable errors in the calculated second-order quantities in Tables II-IV appear to be compatible with the estimated experimental errors and with the magnitude of the contributions of these quantities to the polarizing effect.

The differences between the comparable quantities, in Tables II and III and in Table IV, typically several percent and in some cases larger, are likely to be a consequence of the combined effects of the experimental errors and of the lumping of the second-order piezoelectricity and the second-order permittivity in the components of the cumulative electroelastic tensor. In addition to it one should consider the possible effect of the terms reducing the symmetry of the second-order tensors treated in this paper. These terms were found present in the nonlinear study⁷ of the polarizing effect referred to earlier. It is not at all clear at this stage which of the three mentioned sources of discrepancies is the most serious one.

In general, this work represents a substantial improvement over the former numerical results^{4,5}. It is based on an enlarged set of experimental data covering all the three thickness modes. In addition to the electroelastic effect, it numerically relates the polarizing effect to other second-order phenomena. It yields the values of the computed electroelastic tensor components with substantially reduced errors. Also, the quality of reproduction of the experimental data is improved in spite of their larger quantity. The fact that a part of the experimental data serving as a basis for this paper has already been used⁵ makes it particularly easy to notice the improvement.

This work on the polarizing effect has been prompted by the current interest in second-order phenomena. Its general purpose is to demonstrate that the polarizing effect deserves attention because of the fact that its analysis may contribute to the study of second-order effects.

Acknowledgement

The author appreciates receiving, from Dr. J.A. Kusters, the numerical values for the electroelastic tensor component published only in a graphical form.⁴ He also wishes to thank Mr. G. Ord, Miss R. Shenitzer,

and Miss P. Hruska for their assistance at various stages of preparation of this paper.

This work was supported in part by the National Research Council of Canada under Grant No. A9515, and by York University President's NRC Fund (November 1976) and the Faculty of Arts Minor Research Grant (April 1977).

References

1. K. Hruska, "The influence of a direct electric field on the oscillations of quartz resonators and its dependence on their orientation," *Cs. Cas. Fys.*, vol. A12, pp. 35-38, 1962.
2. K. Hruska, "An attempt at a phenomenological interpretation of the influence of a polarizing field on piezoelectric resonators," *Czech. J. Phys.*, vol. B12, pp. 338-353, 1962.
3. K. Hruska and A. Khogali, "Polarizing effect with alpha-quartz rods and the electroelastic tensor," *IEEE Transactions on Sonics and Ultrasonics*, vol. SU-18, pp. 171-176, July 1971.
4. J.A. Kusters, "The effect of static electric fields on the elastic constants of α -quartz," *Proc. 24th AFCS*, April 1970, pp. 46-54.
5. K. Hruska, "Polarizing effect with piezoelectric plates and second-order effects," *IEEE Transactions on Sonics and Ultrasonics*, vol. SU-18, pp. 1-7, January 1971.
6. K. Hruska, L. Janík and V. Kazda, "The dependence of natural frequency of quartz plates on dc electric field," *Proc. of the Technical University of Liberec, Prague: SPN*, 1968, pp. 27-30.
7. J.C. Baumhauer and H.F. Tiersten, "Nonlinear electroelastic equations for small fields superposed on a bias," *J. Acoust. Soc. Amer.*, vol. 54, pp. 1017-1034, 1973.
8. IRE Standards on Piezoelectric Crystals, *IEEE Standard 176*; also *Proc. IRE*, vol. 37, pp. 1378-1395, December 1949.
9. IRE Standards on Piezoelectric Crystals, *57 IRE 14.S1*; also *Proc. IRE*, vol. 45, pp. 353-358, March 1957.
10. J.F. Nye, *Physical Properties of Crystals*, London: Oxford University, 1957, pp. 116-122, 124, 124-135, 250-251.
11. D.A. Kleinman, "Nonlinear dielectric polarization of optical media," *Phys. Rev.*, vol. 126, pp. 1977-1979, June 1962.
12. K. Hruska and V. Kazda, "The polarizing tensor of the elastic coefficients and moduli for α -quartz," *Czech. J. Phys.*, vol. B18, pp. 500-503, 1968.
13. R. Bechmann and A. Ballato, "The coefficients of second harmonic generation of light (SHG) of noncentrosymmetric crystals," *Research and Development Technical Report ECOM-3529*, May 1972.
14. R. Bechmann, "Elastic and piezoelectric constants of alpha-quartz," *Phys. Rev.*, vol. 110, pp. 1060-1061, June 1958.
15. J.A. Kusters, private communication.

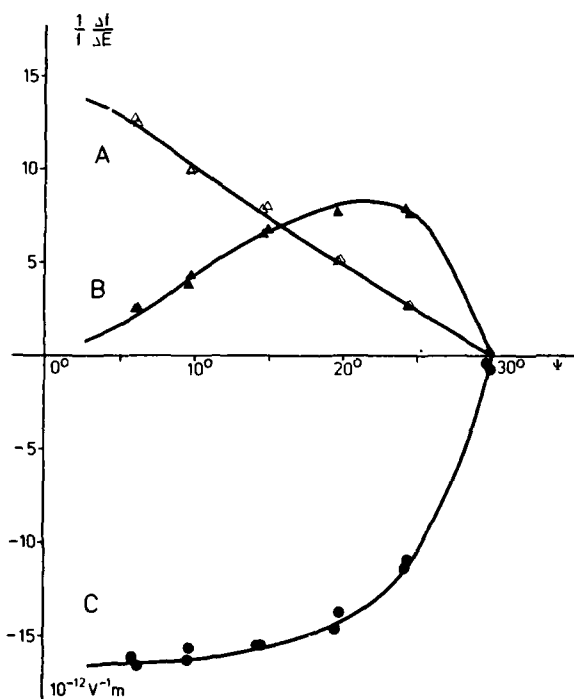


FIGURE 1 The magnitude of the polarizing effect $(1/f) \cdot (\Delta f/\Delta E)$ for the fundamental frequency of the thickness modes of plated cuts of thickness 0.6×10^{-3} m and orientation $(xz1)\psi$. The observed values marked Δ , \triangle , and \bullet correspond to the modes A, B, and C respectively. The curves are calculated by means of Eq. (38) and the second-order quantities from Table II. The appearance of the plotted curves does not change appreciably in this scale when their calculation is repeated using Eq. (39) and the second-order quantities from the column marked $(xz1)\psi$ in Table III.

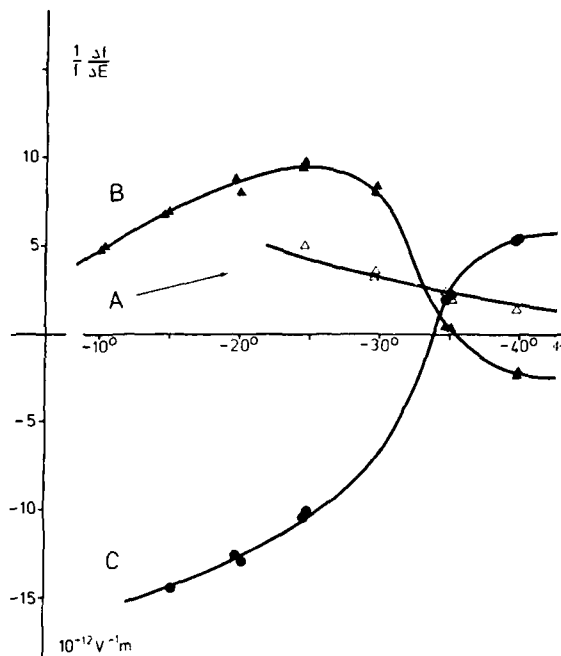


FIGURE 2 The magnitude of the polarizing effect for the plates of orientation $(xzw)\psi$. The curves are calculated by means of Eq. (40) and the second-order quantities from the column marked $(xzw)\psi$ in Table IV. Other explanatory remarks are identical with those made for Figure 1.

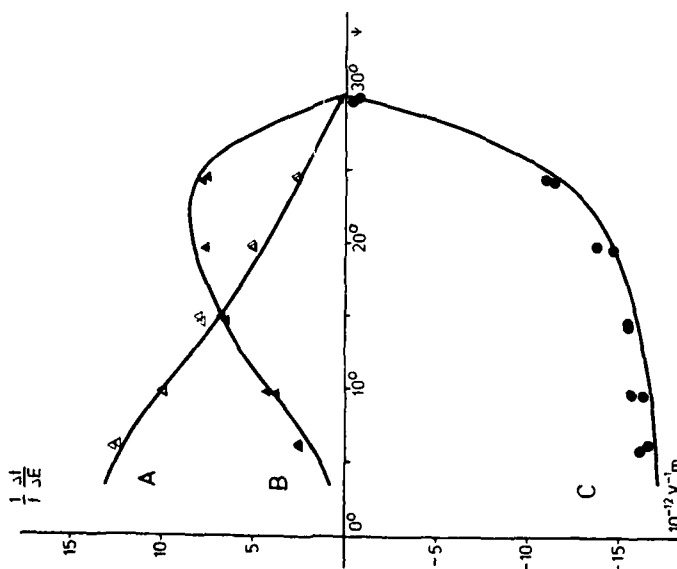


FIGURE 3 The curves predicting the magnitude of the polarizing effect for the plates of orientation $(xz1)\psi$. The plotted points correspond to the observed values. The curves are calculated by means of the observations of the polarizing effect for the plates of orientation $(xzw)\psi$. Other explanatory remarks are identical with those made for Figure 1.

n r s	111	112	113	122	123	321	222	133	223	233	333
	111	121	131	212	231	132	222	313	232	323	333
		211	311	221	312	213		331	322	332	
Δ	1	2	3	4	6	8	9	12	18	27	

TABLE I Correspondence between the three index and the single index notations for the components of the second-order permittivity tensor. A simple assignment rule $\Delta = n . r . s$ generates ten Greek indices Δ easily related in both directions to the original triads of the Latin indices $n r s$. The rule can be used provided that the Kleinman conditions¹¹ are applicable.

	(xz1) ψ
$\delta_{111} + 1.63 \times 10^9 q_{11}$	$NV^{-1}m^{-1}$ 3.06 \pm 0.04
$\delta_{141} - 1.00 \delta_{441} - \delta_{241}$	$NV^{-1}m^{-1}$ 1.06 \pm 0.04
$\delta_{221} + 2.00 \delta_{441} - 1.62 \times 10^9 q_{11}$	$NV^{-1}m^{-1}$ -2.12 \pm 0.08
q_{41}	$10^{-10} NV^{-2}$ -1.14 \pm 0.28
t_1	$10^{-20} FV^{-1}$ 4.27 \pm 1.27

TABLE II Values for some components of the tensors of the second-order effects of α -quartz at 20°C. The values are calculated using the 31 observations of the polarizing effect for the plates of orientation (xz1) ψ shown in Fig. 1. The quantities $\delta_{\lambda\mu n} = \partial c_{\lambda\mu} / \partial E_n$ are the components of the electroelastic tensor of α -quartz. The quantities $q_{\lambda\mu}$ are the components of the second-order piezoelectricity tensor defined as $q_{ijmn} = \partial e_{mij} / \partial E_n$, where the pairs of Latin indices ij and mn , ($i, j, m, n = 1, 2, 3$), are replaced by the Greek λ and μ , ($\lambda, \mu = 1, \dots, 6$) in the usual manner.¹⁰ The quantity $t_1 = \partial \epsilon_{11} / \partial E_1$ is the only independent nonzero component of the second-order permittivity tensor of α -quartz. The coefficients 1.63×10^9 and 1.62×10^9 have the dimensions of Vm^{-1} . The tolerances are stated in terms of the mean probable error.

$\delta_{\lambda\mu n}$	$(xz1)\Psi$	$(xzw)\Phi$
	$NV^{-1}m^{-1}$	
δ_{111}	2.98 ± 0.03	2.90 ± 0.10
$\delta_{141} - 1.00 \delta_{441} - \delta_{241}$	0.93 ± 0.03	0.98 ± 0.09
$\delta_{221} + 2.00 \delta_{441}$	-1.86 ± 0.04	-1.74 ± 0.06

TABLE III Values for some cumulative electroelastic tensor components of α -quartz at 20°C. The values in the two columns are calculated independently, using the observations of the polarizing effect for the plates of orientations $(xz1)\Psi$ and $(xzw)\Phi$ shown in Figures 1 and 2 respectively. These values are obtained if the interpretation of the polarizing effect is based solely on the electroelastic and the converse piezoelectric effects, i.e. neglecting the second-order permittivity and the second-order piezoelectricity. Other explanatory remarks are identical with those made for Table II.

$\delta_{\lambda\mu n}$	$(xzw)\Phi$	$(xzw)\Phi$ $(xz1)\Psi$	$(xz1w)\Psi\Phi$ Kusters
	$NV^{-1}m^{-1}$		
δ_{111}	2.90 ± 0.10	2.96 ± 0.03	2.97
$\delta_{131} - 0.42 \delta_{221} + 0.25 \delta_{341}$	0.66 ± 0.11	0.68 ± 0.05	0.32
$\delta_{141} - 0.26 \delta_{221} - 0.67 \delta_{341}$	0.88 ± 0.04	0.90 ± 0.03	0.39
$\delta_{241} - 0.76 \delta_{221} - 0.67 \delta_{341}$	0.77 ± 0.08	0.88 ± 0.04	-1.46
$\delta_{441} + 0.50 \delta_{221}$	-0.87 ± 0.03	-0.90 ± 0.02	-0.39
$\delta_{153} + 0.60 \delta_{221} - 0.51 \delta_{341}$	-1.03 ± 0.07	-1.01 ± 0.04	-0.78

TABLE IV All components of the cumulative electroelastic tensor that can be calculated from observations of the polarizing effect with the plates of orientation $(xzw)\Phi$. The values calculated using the 29 observations from Fig.2 are given in the column marked $(xzw)\Phi$. The values in the next column are calculated by means of all 61 experimental data available in this paper. The degree of agreement with the values in the last column is properly appreciated only if it is realized that the orders of magnitude of the values for the components δ_{131} , δ_{221} , ..., and δ_{153} given by Kusters^{4,15} are $10 NV^{-1}m^{-1}$ and $10^2 NV^{-1}m^{-1}$. Other explanatory remarks are identical with those made for Tables II and III.

COERCIVE STRESS FOR FERROBIELASTIC TWINNING IN QUARTZ

T. L. Anderson, R. E. Newnham and L. E. Cross
Materials Research Laboratory
Pennsylvania State University
University Park, Pa. 16802

Summary

The coercive stress for ferrobielastictwinning in quartz has been measured using a uniaxial stressing jig at room temperature and at 148, 208, 318 and 355°C. At the three lowest temperatures, stripe-shaped domains were induced with walls parallel to the optic axis. When the stripe-shaped domains spanned the crystal from edge to edge, they began to expand uniformly in the X_2 direction, maintaining the domain walls with minimum strain mismatch. With continued increase in stress, the samples converted completely to the alternate Dauphiné twin state with an audible snapping sound. At room temperature, when the stress was relieved, the sample switched back to its original state, but at 148, 208, 318 and 355°C no switchback was observed.

At temperatures above 318°C, no stripe-shaped twins were observed, but rather an irregular domain expanded in all directions without regard to domain wall restrictions until the entire sample had switched.

In general, as the temperature was increased, the coercive stress for domain nucleation decreased at a rapid rate over the temperature range tested. It appeared as though the structural bias was annealed out at elevated temperatures and that domain wall strain mismatch decreases rapidly with temperature.

Key Words Twinning, Ferrobielasticty, Domains, Dauphiné Twins, Coercive Stress, Mechanical Twinning.

Introduction

Several types of twinning can occur in quartz. 'Brazil' twins are growth twins, and if absent from the original crystal, they cannot be induced later without breaking strong silicon-oxygen bonds. Electrical, or Dauphiné twins on the other hand have orientation states related by a 180° rotation about the trigonal axis; i.e., the positive x-axis in one state becomes the negative x-axis in the complementary twin state (Fig. 1). The atomic shifts necessary to go from one twin state to the other involve motions of less than 0.3Å and require no breaking of silicon-oxygen bonds.

The subtle distortions which occur in Dauphiné twinning are associated with the loss of symmetry which takes place at the β and α phase transition (from point group 622 to 32) and are consistent with the ferroic species 622F32 in the notation of Aizu.¹ For this species it should be possible to move the twin walls by the application of a suitable mechanical stress (ferrobielastict switching) or a suitable combination of elastic and electric fields (ferroelastoelectric switching). In the former case the 'driving force' is associated with the change in sign of the compliance element s_{1123} between twin states, and in the latter case, through the change of sign in the piezoelectric element d_{111} between states.

The phenomenon of stress-induced twinning and detwinning of quartz has been known for many years²; however, this ferrobielastict switching does not appear to have been properly and systematically studied with well defined elastic, thermal and electrical boundary conditions.

Such studies are of basic interest in that they

will contribute to a better understanding of the nature of the bielastict twin wall, its mode of motion and the relation to the phenomena of ferroelectric and ferroelastic twin wall motion. The studies are also of practical importance for three areas:

(i) In attempting to exploit the control of elastic and piezoelectric properties afforded by the twinning operations³, it is important to be able to control the shape, volume and spatial arrangement of the twin components. The elastic coercivity and its orientation dependence under different driving fields are clearly of cardinal importance;

(ii) For the correct processing of present single crystal (monodomain) resonators and control devices, it is obviously highly desirable to know the limiting environmental parameters below which twinning will not be induced;

(iii) Certain geophysical problems require an understanding of the plastic deformation of quartz bearing rocks at very high stress levels. Under these conditions bielastict twin motion may be possible, and may interact with other deformation mechanisms.

In this study, the orientation dependence of the elastic drive function for ferrobielastict switching has been determined for a simple biaxial stress system applied to single domain crystals. The symmetry of the drive function profile has been verified by thermally induced switching experiments under limited laser heating.⁴

For the orientations where the drive field requirements are a minimum, the coercive stress has been examined as a function of temperature using a simple uniaxial stressing jig. In this arrangement the appearance of Dauphiné twins can be monitored optically in polarized light since the photoelastic coefficient π_{1123} changes sign between orientation states.

Experimental

The quartz bars used in the ferrobielastict switching experiments were fashioned from large synthetic single crystals which were commercially produced using hydrothermal techniques. The samples were optically clear and free from obvious imperfections. No twins were detectable in the virgin samples.

X-cut quartz bars were used in the switching experiments. The orientation of each crystal was determined using the Laue x-ray technique. From a single Laue photograph parallel to a diad axis, the X_2 ($= [120]$) and X_3 ($= [001]$) directions as well as the sense (positive or negative) of the X_1 ($= [100]$) axis can be determined.

Having established the orientation of the crystal axes, the samples were cut with a diamond saw into rectangular bars with the faces to be stressed rotated about the X_1 axis to form 45° angles with the X_2 and X_3 axes (Fig. 2). Typical sample dimensions were approximately 0.8 x 0.8 x 0.4 cm³ with the thin dimension parallel to the X_1 axis. The surfaces to be stressed were then lapped parallel and the faces perpendicular to X_1 polished sufficiently to render them transparent.

The pressure cell employed in these experiments is

illustrated in Fig. 3. The two opposing rams were fashioned from cylindrical bars of tungsten carbide with ends polished perpendicular to the ram axis. The sleeve was produced from S.A.E. 1018 mild steel and was bored to produce a ram-sleeve clearance of 0.004 cm. A sample chamber for sample loading and optical observation and two thermocouple openings were machined perpendicular to the sleeve axis.

The stressing jig is illustrated in Fig. 4 and features an Enerpac RC53 hydraulic ram, a B.L.H. Electronics Type C3P1 load cell, and a Daytronic Model 700 digital strain gauge indicator. To stabilize and protect the hydraulic ram and strain gauge from the effects of elevated temperatures, all the pistons are water cooled. All contacting surfaces along the stress path have been machined flat and perpendicular to the stress axis.

The pressure cell was completely encased in a resistance-wound furnace with temperatures controlled by a Wheelco temperature controller and chromel-alumel thermocouples. Sample temperature was also sensed by a second chromel-alumel thermocouple in direct contact with the crystal bar and monitored by a Fluke Model 8300 digital millivoltmeter. Each thermocouple used an ice water reference.

Since the Dauphiné twin states in quartz can be discerned optically under stress in polarized light, a high intensity light source, a polarizer-analyzer combination and a telemicroscope were positioned as illustrated in Fig. 4. An aperture was drilled through opposite sides of the furnace for illumination and observation of the sample.

Results

With a quartz bar loaded in the stressing jig and oriented as indicated in Fig. 2, and with the polarizer and analyzer set to optical extinction, a uniaxial stress was slowly applied. The stress level at which twins were first observed to nucleate was noted and then allowed to increase. Twins continued to nucleate and expand until finally the entire crystal switched suddenly to the alternate twin state.

At room temperature, 23°C, stripe-shaped domains nucleated from the edge of the bars at stress levels between $2.4 - 2.7 \times 10^4$ nt/cm² (Fig. 5a), depending on the sample's past history, and continued to grow in the X₃ direction (Fig. 5b).

The advance of these domains was not always continuous but was often abrupt and step-like. With the application of still higher pressures, the domains eventually spanned the crystal from edge to edge; new states continued to nucleate and grow until the samples were completely lined with domains. At this point domains began to expand in the X₂ direction (Fig. 5c) until finally the sample quite abruptly switched to its energetically lower twin state. Switch-over was accompanied by an audible snap.

At room temperature, the stress required for switch-over was $4.0 - 5.5 \times 10^4$ nt/cm², once again depending on the crystal's past history. Continued increases in the applied stress beyond the point of switch-over had no further visible effect on the twin state of the crystal and no additional twin motion was observed.

When the pressure was slowly removed from samples which were partially twinned, a switch-back process occurred at stresses of about 2.3×10^3 nt/cm². The shape and orientation of the domain states associated with switch-back are quite different from the well

defined, highly oriented, stripe-shaped domain states observed in the forward process. Switch-back began in the area of the crystal which was last to twin in the forward process. The domain was singular and shapeless and expanded on a broad hazy front until the entire crystal had switched back at zero applied stress.

It was observed that the growth of domain states proceeds with a definite velocity and that there is a lag between applied stress and domain growth. Once the twinning process has begun, the applied stress can be held constant while twins continue to grow for some time thereafter. Typically, for virgin material, stripes grow in length parallel to X₃ until they span the crystal from edge to edge and then begin to grow wider along the X₂ direction.

As indicated previously, the initial coercive stress and subsequent switch-over stress depend on the past history of the sample. The sample with the lowest coercive stress contained several twins prior to performing the stressing process. The samples with the largest coercive stress had previously been squeezed in a direction which stabilized the existing twin state.

With a second application of stress in the forward direction, it was discovered that the coercive stress to nucleate twins was typically an order of magnitude lower than that measured during the first stress application. Likewise, the stress required for complete switch-over was reduced to about half the value recorded for the first switch-over. Twins were observed to begin nucleation in the range $3.4 - 4.4 \times 10^3$ nt/cm², with complete switch-over occurring at $1.5 - 2.9 \times 10^4$ nt/cm². The switch-back process became noticeable at 1.8×10^3 nt/cm² as the applied stress was relaxed.

Subsequent stress applications showed that after the second switch-over, the pressure necessary to initiate domain movement and switch over stabilized at slightly lower values. If the pressure was increased beyond that required for switch-over, the stress necessary to move twins in the subsequent trial, as well as that required for switch-over, was reduced. Repetition of this process twenty times at stress levels well beyond those required for switch over (up to 9×10^4 nt/cm²) caused the crystal to cease switching back; the crystal remained in the alternate Dauphiné twin state.

Rotating the crystal 90° about the X₁ axis and once again applying a stress caused the sample to switch back to the original state. The stress level necessary to accomplish this reversal was much lower than that which was necessary to go in the forward direction. Once the crystal has undergone switch-back to the original state, it does not experience any back switching and remains in the original state.

If once again the crystal is rotated 90° about the X₁ axis, back to the initial stressing orientation, one notes that the shape of the stress induced twins has changed. Instead of being stripes which grow along the X₃ direction, the domains are pointed with an apex angle of 90° and grow equally fast in both the X₂ and X₃ directions. After switch-over, no switch-back is observed.

That the sample had indeed switched from one twin state to the other during these experiments was also verified electrically. The sign of the piezoelectric modulus d_{111} was measured prior to stressing the crystal. Measurement after the crystal had been observed to optically switch showed that the d_{111} coefficient had changed sign, indicating that the polarity of the X₁ axis had changed and that the crystal had truly switched.

At temperatures of 148°C and 208°C, the results for two different samples were similar to those obtained at room temperature. Stripe-shaped twins nucleated at an edge of the crystal and expanded in the X_3 direction until they extended from edge to edge. At that point, growth in the X_2 direction was observed until finally the crystals switched over to the alternate twin state. No switch-back was observed when the stress was removed. This also was verified by measurement of the sign of the piezoelectric modulus d_{111} .

The only detectable difference in the results recorded at 148°C and 208°C was a lowering of the stresses required to nucleate twins and for complete switch-over. At 148°C, twins began to appear at 1.20×10^4 nt/cm² and complete switch-over was evident at 1.80×10^4 nt/cm². At 208°C, twins began growing at 1.12×10^4 nt/cm² and complete switch-over was observed at 1.93×10^4 nt/cm².

At 318°C, twinning was observed to begin at 3.40×10^3 nt/cm² and the sample had completely switched over at 1.28×10^4 nt/cm². However, stripe-shaped domains no longer were evident. The domains bear a strong resemblance to those described in the switch-back process. The domain began at an edge and was singular. As stress was increased, the domain grew as a volume with irregular shape and orientation and moved continuously across the sample until complete switch-over resulted. There was no audible snap at switch-over and no switch-back when the stress was released.

Results obtained at 355°C showed twins which behaved similarly to those noted at 318°C. Domains were initiated at a stress level of 2.27×10^3 nt/cm² with complete switch-over occurring at 4.5×10^3 nt/cm². It is interesting to note that the coercive stress has been reduced, at these higher temperatures, to the point where, but for the temperature, a good hard pinch with the fingers would be sufficient to induce twinning. Clearly, this possibility should be kept in mind during any high temperature processing of quartz for resonator applications.

A comment should be made concerning the size of the stripe-shaped domains. While no concerted effort has yet been made to measure the size of each domain, an estimation of the density of twins across the face of the sample would seem to indicate that the domain width is between 2 - 20 μ m. Since no judgement was made concerning the depth of the field of view, it is not known how deep the domains extend into the bulk of the sample. It was observed, however, that domains appear to nucleate and grow in areas which are behind existing observed domains and hence the twins do not extend continuously from surface to surface along the polar axis.

Discussion

To explain the observed domain patterns in quartz we examine the free energy function for twinned crystals.⁵ Ferroelastic switching is governed by the term $\Delta s_{ijkl} \sigma_{ij} \sigma_{kl} / 2$ involving elastic compliances (s_{ijkl}) and applied mechanical stresses (σ_{ij}, σ_{kl}). α -quartz has six independent elastic compliance coefficients: $s_{1111}, s_{1122}, s_{1133}, s_{1123}, s_{2323}$, and s_{3333} . For a composite crystal containing Dauphiné twins referred to a common set of axes, the sign of the coefficient s_{1123} is reversed for the two twin states. The other five independent coefficients are not affected by Dauphiné twinning. The active coefficient s_{1123} is symmetry-related to a number of others: $s_{1123} = s_{1132} = s_{2311} = s_{3211} = -s_{2223} = -s_{2232} = -s_{2322} = -s_{2222} = s_{1312} = s_{3112} = s_{3121} = s_{1213} = s_{1321} = s_{1231} = s_{2113} = s_{2131}$. Substituting these

terms in $\Delta s_{ijkl} \sigma_{ij} \sigma_{kl} / 2$ and simplifying, we obtain $4s_{1123}(\sigma_{11}\sigma_{23} - \sigma_{22}\sigma_{23} + 2\sigma_{12}\sigma_{13})$ as the energy difference between Dauphiné twin states.

Thus the stress combination required to produce ferroelastic switching is $\sigma_{11}\sigma_{23} - \sigma_{22}\sigma_{23} + 2\sigma_{12}\sigma_{13}$. Whenever this combination is large, twin wall motion may result. This expression has been evaluated for X-cut crystals assuming that the stresses lie within the plane of the plate.

The coordinate system for X-cut quartz is shown in Fig. 6. Let σ be a compressive stress along X'_3 , an arbitrarily chosen reference direction in the plane of the plate. X'_3 forms an angle α with respect to X_3 and an angle $\pi/2 - \alpha$ with X_2 , where X_2 is the crystallographic [120] direction and X_3 is [001]. X_1 , the direction normal to the X-cut plate, is [100]. Transforming the stress σ into stress components σ_{ij} , we find that the switching stress combination $\sigma_{11}\sigma_{23} - \sigma_{22}\sigma_{23} + 2\sigma_{12}\sigma_{13}$ is equal to $-\sigma^2 \sin^3 \alpha \cos \alpha$. From this result it is apparent that stresses along X_3 ($\alpha = 0^\circ$) and X_2 ($\alpha = 90^\circ$) are useless in promoting twin wall motion. The best orientation is $\alpha = 60^\circ$, as shown in Fig. 6, where $\sigma_{11}\sigma_{23} - \sigma_{22}\sigma_{23} + 2\sigma_{12}\sigma_{13}$ is plotted as a function of stress orientation. The solid line shows the stress directions stabilizing one twin, the dashed line the other. This means that compressive stresses in the first and third quadrants favor one twin, while orientations in the second and fourth favor the other. Thus, if the plate is untwinned initially, only odd-quadrant stresses will produce switching while even quadrant stresses further stabilize the existing twin state. These calculations are confirmed by the experimental observations.

The stress angle $\alpha = 45^\circ$ was chosen to facilitate the crystal being stressed in both the forward and reverse directions without having large differences in the stress levels. With $\alpha = 45^\circ$, the stress required to produce domain wall motion is increased only 23% from that which is required for stresses oriented along the easiest direction ($\alpha = 60^\circ$). Had $\alpha = 60^\circ$ been chosen as the initial stress direction, the stress required for domain reversal ($\alpha = -30^\circ$) would have been increased by 67%, possibly producing plastic deformations and irreversible damage to the crystal which would not have been experienced in the initial orientation. Thus, $\alpha = 45^\circ$ is the most useful stress direction for the geometry utilized in this experiment.

It was observed that the bielastic domain walls in quartz prefer to develop parallel to the X_3 optic axis. This domain wall preference can be explained in the following manner. Under stress, the domains in a ferroelastic react differently because of differences in the elastic moduli. For a given set of stresses σ_{kl} , the induced strain in the first orientation state is $\epsilon_{ij} = s_{ijkl} \sigma_{kl}$. In the second orientation state, the same set of stresses produce strains $\epsilon'_{ij} = s'_{ijkl} \sigma_{kl}$. The strain mismatch $\Delta \epsilon_{ij}$ between domains is equal to $\epsilon_{ij} - \epsilon'_{ij} = (s_{ijkl} - s'_{ijkl}) \sigma_{kl}$. All mismatch disappears when the applied stress is zero. It is also zero when the difference in compliance is zero; however, this can only be true for certain directions in a ferroelastic material. In quartz the strain mismatch between Dauphiné twins is:

$$\begin{aligned} \Delta \epsilon_{11} &= 4 s_{1123} \sigma_{23} \\ \Delta \epsilon_{22} &= -4 s_{1123} \sigma_{23} \\ \Delta \epsilon_{33} &= 0 \\ \Delta \epsilon_{23} &= 4 s_{1123} (\sigma_{11} - \sigma_{22}) \\ \Delta \epsilon_{31} &= 8 s_{1123} \sigma_{12} \\ \Delta \epsilon_{12} &= 8 s_{1123} \sigma_{31} \end{aligned}$$

where $s_{1123} = -4.52 \times 10^{-8} \text{ cm}^2/\text{nt}$.

As the equations show, strain mismatch depends on applied stresses but is always zero for certain directions such as X_3 . If an axial system identical to that of Fig. 6 is chosen, the angular dependence of the strain mismatch in the $X_2 - X_3$ plane can be evaluated for the stress geometry used in this study. The result is that a strain compatible direction other than along X_3 exists and is dependent on the stress angle α . For the stress angle $\alpha = 45^\circ$, the direction of zero stress mismatch is at an angle -64.43° in the $X_2 - X_3$ plane. Nevertheless, no domains were observed with walls parallel to this direction.

Since all strain differences are a function of the elastic compliance element s_{1123} and since s_{1123} decreases with temperature to zero at the α - β phase transition at 573°C , the induced strain incompatibilities have virtually disappeared at temperatures near the transition. The experimental results at 308°C indicate that the compliance element s_{1123} has been sufficiently reduced to allow domain walls to nucleate and grow in nearly any direction.

Recall that the bielastic energy difference between the two Dauphiné twin states is also proportional to s_{1123} . However, since the energy difference is quadratic in stress while the strain mismatch is linear, the strain mismatch evidently relaxes before any pronounced reductions in bielastic switching efficiencies are encountered. Even though s_{1123} approaches zero at the 573°C transition, the energy barrier which must be overcome to drive one domain state into its twin state also decreases and at a faster rate. This is supported by the observation that the stress required to initiate twinning decreases with increasing temperature. Spontaneous Dauphiné twinning on the microscale several degrees below the 573°C transition has been reported for many varieties of quartz.⁶

At lower temperatures the twins expand with the preferred wall direction parallel to X_3 . When the twins extend from edge to edge, they begin to grow in the X_2 direction and hence wall compatibility is retained.

The phenomenon of spontaneous switch-back at room temperature is probably due, at least in part, to the fact that the crystals were produced hydrothermally below the phase transition. A structural bias was probably established when impurities and dislocations were trapped by the lattice in such a way as to pin or stabilize the original structural state. These serve as reservoirs of residual internal strains and when all external forces are relaxed, these strains give rise to switch-back. At elevated temperatures, the strains are evidently annealed out as the sample is able to remain in the complementary twin state without undergoing back switching.

References

1. K. Aizu, J. Phys. Soc. Japan **34**, 121 (1973).
2. W. A. Wooster, N. Wooster, J. L. Rycroft and L. A. Thomas, Proc. Inst. Electr. Engn. (Great Britain) **94**, 927 (1947).
3. R. E. Newnham, C. S. Miller, L. E. Cross and T. W. Cline, Phys. Stat. Sol. (a) **32**, 69 (1975).
4. T. L. Anderson, R. E. Newnham, L. E. Cross and J. W. Laughner, Phys. Stat. Sol. (a) **37**, 235 (1976).
5. R. E. Newnham and L. E. Cross, Mat. Res. Bull. **9**, 1021 (1974).
6. R. A. Young, Final Report for Contract No. AF 49 (638)-624 (unpublished).

Acknowledgements

This study was sponsored by the National Science Foundation through Grants ENG-75-02398 and DMR-74-21596. We also wish to thank P. Ryll, R. Hendrickson, R. Laudise, E. Kolb, F. Welsh, N. Coda, J. Balascio, and N. Lias for supplying the quartz crystals. The advice and assistance of our colleagues at Penn State is acknowledged with gratitude.

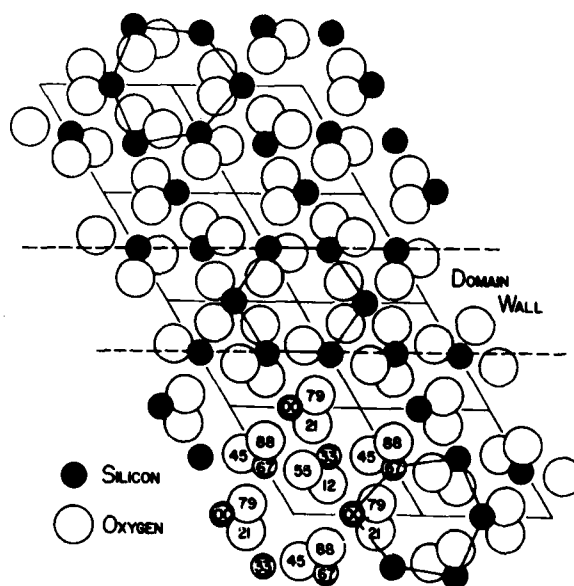


Figure 1: The atomic structure of quartz in the region of a Dauphiné twin wall. The wall is pictured as being one unit cell thick, although there is no direct evidence of this. Atom heights are expressed in units of 10^{-2} of a cell.

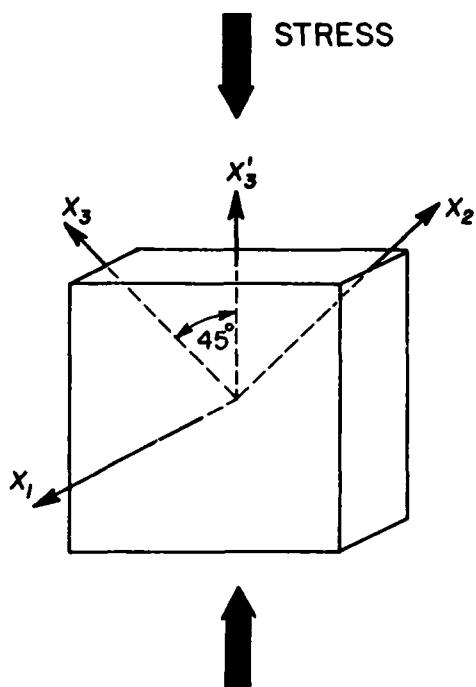


Figure 2: The sample and stress orientations with respect to the crystal axes. Typical sample dimensions are $0.4 \times 0.8 \times 0.8 \text{ cm}^3$ with the thin dimensions parallel to the X_1 axis.

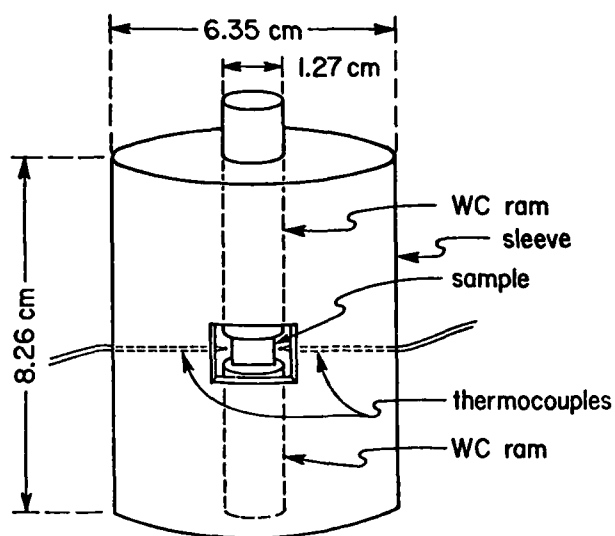


Figure 3: The pressure cell employed in these experiments featuring tungsten carbide rams and a sample chamber for sample loading and optical observation.

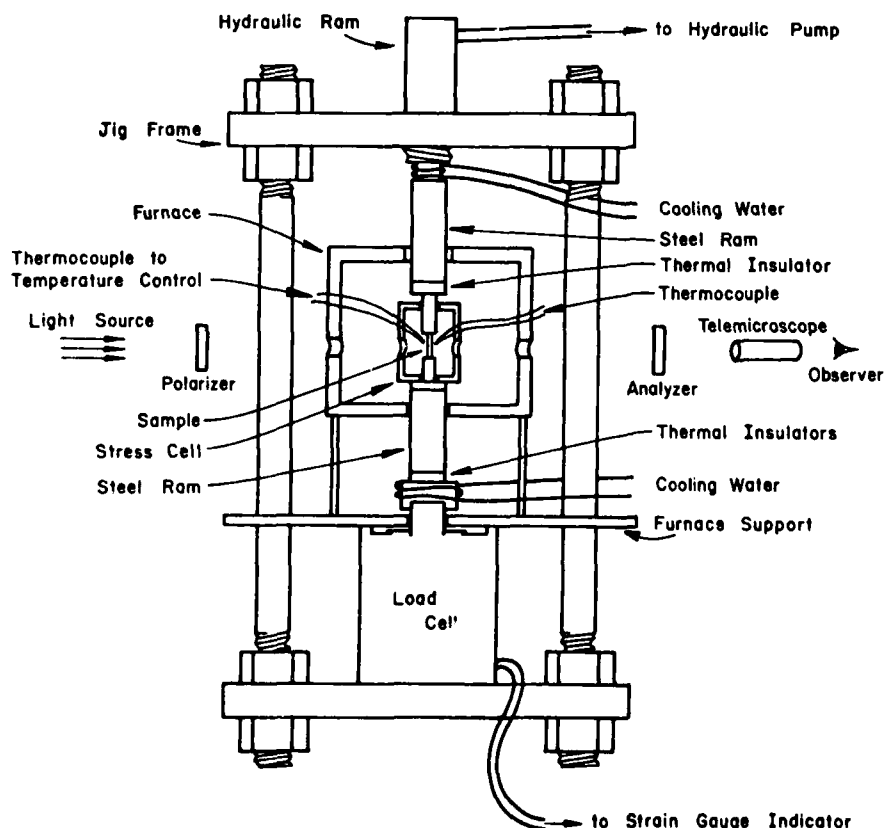


Figure 4: The stressing jig showing the positioning of the pressure cell, hydraulic ram, furnace and load cell. Indicated also is the optical system used to observe the domains under stress.



(a)

(b)



(c)

Figure 5: Mechanically-induced ferroelastic twins in quartz as observed under stress in polarized light. At temperatures up to 208°C stripe-shaped twins nucleated at an edge (a) and expanded in the X_3 direction; (b) with increasing stress until they extended from edge to edge. At that point, growth in the X_2 direction and (c) was observed until the crystal was completely lined with twins. Additional force then caused the crystal to completely switch over to the alternate twin state.

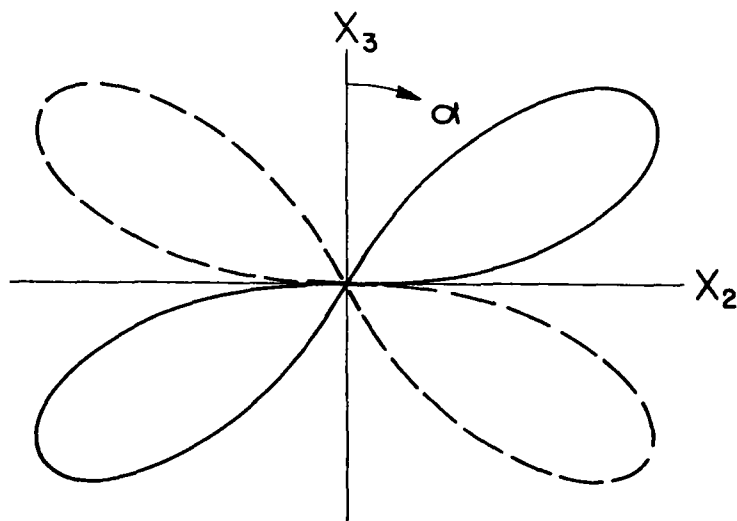
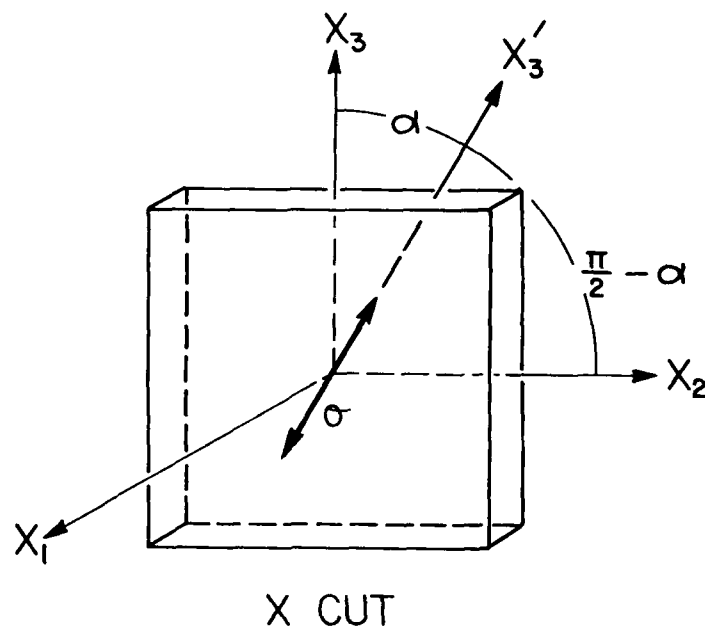


Figure 6: The upper figure illustrates the coordinate system used to derive the switching stress combination for an X-cut quartz plate X_3 represents the stress (σ) direction and lies in the (X_2, X_3) plane; the angle α defines the angle formed between the stress direction and the X_3 -axis. In the lower figure is plotted the switching stress combination ($\sigma^2 \sin^3 \alpha \cos \alpha$) as a function of the angle α . The solid lines indicate directions stabilizing one Dauphiné twin state, while the dashed line represents directions stabilizing the other twin state. Notice that stresses directed along either the X_2 - or X_3 -axis will be ineffective. The directions of maximum stabilization are $\alpha = 60^\circ$ for one twin state and $\alpha = 120^\circ$ for the other.

HYDROTHERMAL SYNTHESIS OF ALUMINUM ORTHOPHOSPHATE

Ernest D. Kolb and Robert A. Laudise
Bell Laboratories
Murray Hill, New Jersey 07974

Summary

We have investigated extensions and modifications of previous AlPO_4 growth techniques, designed to overcome the difficulties posed by the fact that the temperature coefficient of solubility of AlPO_4 is negative. On the assumption that solubility could be improved even at higher temperatures provided pressure was kept high, we investigated crystal growth in the range 150–325°C at 1,000–10,000 psi. The preparation of nutrient material of large enough particle size that it would not be convected in the hydrothermal fluid and become included in the crystal posed a problem. Successful techniques for the recrystallization of small particle AlPO_4 by slowly raising the temperature of a hydrothermal vessel are described. A process for growth by slowly increasing the temperature of a saturated solution and a process for growth in a temperature gradient where the large particle nutrient is contained in a basket in the upper cooler region of the autoclave are described. These processes have been used to prepare crystals grown on seeds at rates up to 16 mil/da.

Introduction

Aluminum orthophosphate AlPO_4 , lithium tantalate, LiTaO_3 , and α quartz, SiO_2 , are the only piezoelectric materials where a suitably-rotated plate of single crystal material can provide an oscillator with a very small frequency-temperature coefficient. Both LiTaO_3 and SiO_2 have been extensively studied and evaluated and both find piezoelectric application. The difficulty of preparing single crystals of AlPO_4 and its comparatively rare occurrence as reasonably-sized crystals (as the mineral Berlinite³) in nature have inhibited its investigation and use as single crystal material. The preparation of J. R. Stanley³ and the unpublished work of W. R. McBride⁴ where Stanley's procedures were in general followed, were the only crystal growth studies known to us when we began this work.

We report here extensions and modifications of Stanley's technique designed to overcome the difficulties posed by the fact that the temperature coefficient of solubility of AlPO_4 is negative. Stanley's data³ indicates solubilities of 3m/liter in 6.1m (molar) H_3PO_4 at 145°C and ~100 psi and 0.3m/l at 245°C and ~500 psi. Solubilities at ambient pressure are lower and at temperatures below 150°C $\text{AlPO}_4 \cdot 2\text{H}_2\text{O}$ ⁵ and other hydrates are stable. AlPO_4 is isoelectronic and isostructural with α SiO_2 and undergoes a phase change at ~584° to the tridymite form, so that a solution growth technique is indicated. Although flux growth might be considered, in view of the success in growing large crystals achieved with other materials we viewed hydrothermal growth as most promising. Stanley used pyrex glass tubes as containers and grew his best crystals by introducing seeds into solutions of $\text{NaAlO}_2 + \text{H}_3\text{PO}_4$ reacted to form AlPO_4 . Growth had to be interrupted every two days to replenish the solution. The temperature was held constant at 165°C during the two-day growth cycle over typical total periods of 50 days. In addition, Stanley grew crystals by slowly increasing

temperatures (1°C every second day from 133° to 155°C). We felt that improved kinetics of crystallization might result at higher temperatures and that the decrease in solubility at higher temperatures might be overcome by raising the pressure. Thus, we decided to investigate crystal growth in the range 150° – 340°C at pressures from ~1,000 – 10,000 psi. In addition, techniques for continuous growth in a temperature gradient and for the preparation of nutrient were studied.

Experimental

Pt. lined Mory⁷-type autoclaves were used since the solvent, aqueous H_3PO_4 , suggested from Stanley's work was felt to be reactive enough to require a noble metal system. Seed frames, nutrient baskets, wire to support seeds, etc., were also Pt. Furnaces, external thermocouples, hot plate, heaters and methods of temperature measurement and control were similar to those previously described⁸ for the hydrothermal growth of other materials so that only a brief outline need be given here.

Autoclaves were 6.93" length x 1.25" diameter (internal dimensions). They were placed on stainless steel hot plates heated by 900 watt Chromalox^{*} strip heaters. Temperature distribution along the autoclave was regulated by adjusting the depth of micaceous insulation in which it was buried. In addition, temperature distribution along the autoclave was regulated by the use of additional 300 watt Watlow[†] cylindrical band heaters strapped directly to the autoclave. The furnace assembly was surrounded by fire brick and shielded by steel plates. Temperature sensing and control was by means of chromel-alumel thermocouples positioned along the autoclave. The temperature controllers were L and N Electromax II⁺⁺ proportional controllers which permitted either constant temperature or slow temperature increase experiments. Temperatures as reported herein are for externally strapped thermocouples but are probably within $\pm 5^\circ\text{C}$ representative of internal temperatures. Pressures reported are those from Kennedy's⁹ p-V-T data for pure water since no equations of state in the H_3PO_4 system have been reported.

RESULTS AND DISCUSSION

Preparation of Nutrient and Seeds

Small particle commercially available AlPO_4 (ALFA Crystal^{**} and ROC/RJC^{***}) is variable in stoichiometry, ranging from Al-10-20%; P 7-16% (AlPO_4 theoretical Al 22.12%, P 25.39%). Small

^{*}Edwin L. Wiegand Co., Pittsburg, Pa. 15208

^{**}Alfa Products, Ventron Corp., P.O. Box 209, 162 Andover Street, Danvers, Mass. 01923

^{***}Research Organic/Inorganic Chemical Corp., 507-519 Main Street, Bellville, N. J. 07109 (neither of the materials gives x-ray powder patterns for AlPO_4 ; they can only be viewed as convenient sources of Al and P in soluble form)

[†]Watlow Electric Mfg. Co., 1376 Ferguson Ave., St. Louis, Mo. 63141

⁺⁺Leeds and Northrup, North Wales, Pa., 19454

particle nutrient can be swept along by the strong convection in hydrothermal solutions and included in grown crystals so that larger particle nutrient is desirable. A few grown crystals obtained from W. R. McBride were cut up using a stainless steel wire saw and 600 grit SiC - H₂O slurry but most seeds were prepared by building up spontaneously nucleated crystals in several runs. Thus, a technique for large particle nutrient and large spontaneously nucleated crystal preparation was needed.

As mentioned above, the temperature coefficient of solubility of AlPO₄ is likely to be negative in the temperature-pressure range of interest. Thus, in any sort of temperature gradient (where the bottom of the autoclave is hotter) AlPO₄ would be expected to preferably nucleate in the bottom region. In the usual hydrothermal geometry, the bottom of the autoclave is hotter and the convection provides mixing so that growth rates are never transport limited (except for diffusion close to growing crystal interfaces). It would be preferable to maintain this geometry for AlPO₄, placing nutrient in the upper (cooler) region of the autoclave. To do so, requires nutrient of particle size capable of being retained in a container such as a relatively open mesh basket which does not severely restrict solvent access to the nutrient. This is an additional reason for desiring large particle size nutrient and for the necessity of nutrient preparation. Another important reason is to assure stoichiometric nutrient for subsequent reproducible growth studies. Thus, we decided to conduct nutrient preparation placing the starting powder in the bottom of the autoclave and maintaining either isothermal conditions or small temperature gradients along the autoclave. These conditions were obtained by placing the autoclave on a fire brick base and heating with three equally spaced band heaters.

In initial experiments, fine powder commercial AlPO₄ was placed in the bottom of Pt-lined autoclaves and the autoclaves were filled to 80% of their free volume with 6.1M H₃PO₄. The autoclaves were rapidly heated to 150°C and then steadily raised to 275° at a rate of 10°/day. The temperature difference between bottom and top of the autoclave was 2-3° with the bottom hotter.

The amount of AlPO₄ added was believed to produce an undersaturated solution at the beginning of the heating cycle. The product obtained greatly increased in particle size and in all cases gave X-ray and chemical analysis for stoichiometric AlPO₄. Usually a number of 3-4mm spontaneously nucleated crystals were attached to the autoclave wall. Due to the retrograde solubility of AlPO₄, it was necessary in both nutrient preparation and seed growth runs to quench the autoclave in cold water, from the final operating temperature to room temperature, and to remove the products immediately so as to prevent redissolving. A series of runs to explore conditions for maximum yield of large particle product was made, some of which are summarized in Table 1.

Yield in Table 1 is reported as yield of product greater than 60 mesh and as yield of product less than 60 mesh, since only AlPO₄ with a particle size >60 mesh was found appropriate to be contained in baskets for later crystal growth studies.

In the initial stage, the conditions of R2262 were considered standard. Temperature was increased manually in a stepwise manner daily. A number of runs (not shown in Table 1) were made at these conditions and the yields of nutrient >60 mesh were in all cases from 9-11 gms. Fill was varied from 80-83% with no

TABLE 1

Run No.	2248	2250	2262	2279	2292
Solvent H ₃ PO ₄ (Molar)	6.1	6.1	6.1	6.1	6.1
% Fill	82	82	82	80	81
AlPO ₄ charge-gms	36	36	36	36	45 presat. @ 25°
Initial Temp.°C	150	150	151	150	150
Final Temp.°C	300	250	274	278	275
Temp. Diff. (ΔT)°C	2-3	2-3	2-3	3-5	1-2
Heating rate°C/day	15	5	10	10	10
Method of heat rate incr.	Manual*	Manual*	Motor	Motor	Motor
Yield in gms. > 60 mesh	12.6	14.1	10.7	17.6	19.5
< 60 mesh	13.4	8.1	10.6	3.8	8.1
Total yield	26.0	22.2	21.3	21.4	27.6
Run Duration in da.	13	23	14	15	15

* manual step wise daily

** 60 mesh - U.S. sieve size = 0.25mm

significant effect on yield of >60 mesh nutrient. Temperature difference (ΔT) was increased to 20°C with a reduction of yield, so that essentially isothermal conditions (ΔT's only a few degrees) were judged best. Experiments with a negative ΔT (bottom of autoclave colder than top) were attempted even though it was felt that in the absence of convection growth would be small. A ΔT of (-) ~20° produced a low yield (~5 gms) of >60 mesh nutrient. The yield was not strongly dependent on fill between 80 and 82%, although there was a slight reduction in preliminary experiments at 63%. In R2250, the heating rate was halved, the yield >60 mesh was essentially unchanged with the result that the yield/unit time of the experiment was greatly reduced so that reduced heating rate experiments appeared unprofitable.

In R2248, both the heating rate and the final temperature were increased with an obvious increase in yield >60 mesh and an increase in yield >60 mesh/unit time of the experiment. A series of runs were made where the temperature was increased continuously with an appropriate clock-driven mechanism. Run 2279 is representative of these experiments and as can be seen the yield >60 mesh is improved.

In a series of runs in which R2292 is representative, the H₃PO₄ solution was presaturated at room temperature with AlPO₄. As can be seen, a possible slight increase in yield of >60 mesh product results.

From the above, the best conditions for nutrient and spontaneously-grown seeds were chosen to be the conditions of R2292. In most of the above runs a few larger (3-4mm) spontaneously nucleated crystals suitable for preliminary seeds were produced. Figure 1 shows spontaneously nucleated seeds on the left and >60 mesh nutrient on the right.

Growth on Seeds

Spontaneously nucleated seeds and oriented seeds cut from grown crystals were mounted on a Pt frame which was placed in the bottom (hotter) region of an

AD-A086 221

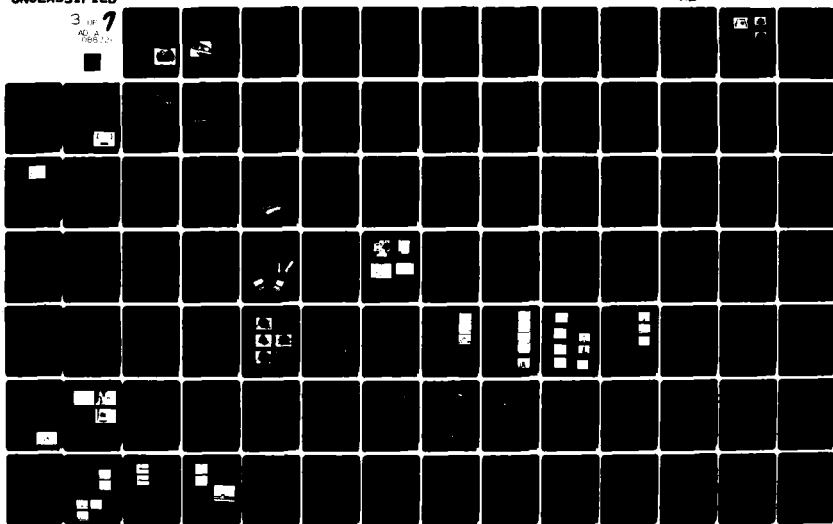
ARMY ELECTRONICS COMMAND FORT MONMOUTH NJ
PROCEEDINGS OF THE ANNUAL FREQUENCY CONTROL SYMPOSIUM (31ST) HE-ETC(U)
1977

F/6 9/5

UNCLASSIFIED

NL

3 OF 7
AD-A086 221



A
82

autoclave and nutrient (>60 mesh particle size) prepared as described above was placed above in a Pt basket (Figure 2) in the upper region. Thus, the seeds were in a hotter supersaturated region of the autoclave and the temperature gradient (hot on bottom) was still arranged so as to allow convection.

In experiments, where the temperature difference was more than $\sim 3^\circ$ (nominally isothermal conditions) the autoclaves were placed on a hot plate and heated from below with ΔT adjusted with one band heater in the top region. The isothermal experiments were conducted with the heater configuration previously described.

Growth by Slow Heating

Using 6.1m H_3PO_4 , ~ 20 gms of nutrient always >60 mesh, a series of runs was made where the temperature was increased continuously. The preliminary optimum conditions for growth were found to be 80-82% fill with a temperature increase of $2^\circ F/da$ from 150 to $180^\circ C$. The growth rates were determined by measuring the increase in dimensions along the c axis of crystals which had already capped with (0111) faces. Rates as high as 16 mil/da were observed. Geometric considerations allow the rate on a (0111) seed to be calculated as 16 mil/da $\cos \theta$ where $\theta = 68^\circ 41'$ the angle between (0111) and the (0001), thus the (0111) rate is ~ 6 mil/da.

Temperature Differential (ΔT) Growth

Using 6.1m H_3PO_4 , ~ 20 gms of >60 mesh nutrient, a series of seeded runs were made to effect growth where supersaturation was provided by a temperature differential. Fills between 70 and 83% were investigated. Growth rates were reduced at 83% and seeds tended to be lost by dissolution at 70% so that 80% fill seemed optimum. Growth temperatures between 200° and $300^\circ C$ were used. ΔT 's were 50° except in one $25^\circ \Delta T$ experiment where the seeds dissolved during the warmup. Growth rates varied between 3 and 7 mils/da at 250 and 300° and were much reduced at 200° . Thus, in preliminary ΔT experiments growth rates approaching the useful range have been obtained under conditions which should provide continuous growth.

From the above, we can see that both growth by a temperature increasing method and growth by a temperature gradient (ΔT) method can be used to prepare $AlPO_4$ single crystals. The advantage of the ΔT method is that the amount of material to be deposited is not limited to the amount present in solution. Typical grown crystals are shown in Figure 3.

Conclusions

Techniques for the preparation of large particle $AlPO_4$ nutrient have been developed as have techniques for preparing spontaneous nucleated seeds for preliminary growth experiments. Seeded growth both by (1) slowly warming a saturated solution and by (2) transport from cooler large particle nutrient contained in a basket in the upper region of an autoclave, to seeds in the lower region have proven feasible. Growth rates as high as 16 mil/da have been obtained. Thus, processes for the growth of $AlPO_4$ by hydrothermal means, including continuous growth without need for replenishment of solution seem practical. The visual quality (when seed quality is good and mounting wires do not interfere) can be good.

Acknowledgments

We gratefully acknowledge the receipt of $AlPO_4$ seeds crystals from Dr. John Vig, U. S. Army Electronics Command, Ft. Monmouth, N. J. and Dr. W. R. McBride of the Naval Weapons Center, China Lake California. Also, we acknowledge the able technical assistance of A. J. Caporaso who conducted the hydrothermal experiments. We also thank R. L. Barns for help with seed orientation.

References

1. Y. P. Chang, G. R. Barsch, IEEE Trans. on Sonics and Ultrasonics **23** (2), 127, March 1976 see also A. Jhunjhunwala, J. F. Velelino and J. C. Field, J. Appl. Phys., **48** (3) 887, (1977).
2. The System of Mineralogy, C. Palache, H. Berman and C. Frondel, Wiley, N. Y. Vol. II p. 696 (1970).
3. J. M. Stanley, Ind. Engr. Chem. **46**, 1684 (1954).
4. W. R. McBride, Dept. of Navy Weapons Center, China Lake, California, private communication.
5. E. Z. Arlidge, V. C. Farmer, B. D. Mitchell, and W. A. Mitchell, J. Appl. Chem. **13**, 197 (Jan.)1973.
6. L. H. Cohen, W. Klement, Jr., Amer. Mineral. **58**, 796, (1973).
7. G. W. Morey and P. Niggli, J. Amer. Ceram. Soc. **35**, 1086, (1952).
8. The Growth of Single crystals, R. A. Laudise, Prentice Hall, Inc., Sec. 7.3 p. 275ff (1970).
9. G. C. Kennedy, Amer. J. Sci. **248**, 540 (1950).



Figure 1 - Recrystallized $AlPO_4$ nutrient > 60 mesh and larger spontaneously nucleated crystals.

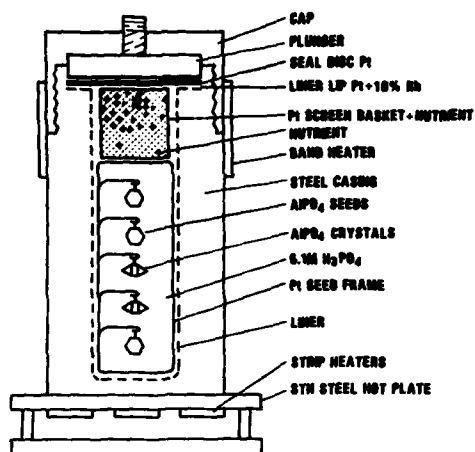


Figure 2 - Schematic of platinum-lined Morey vessel showing Pt basket, frame, nutrient, seeds and crystals.

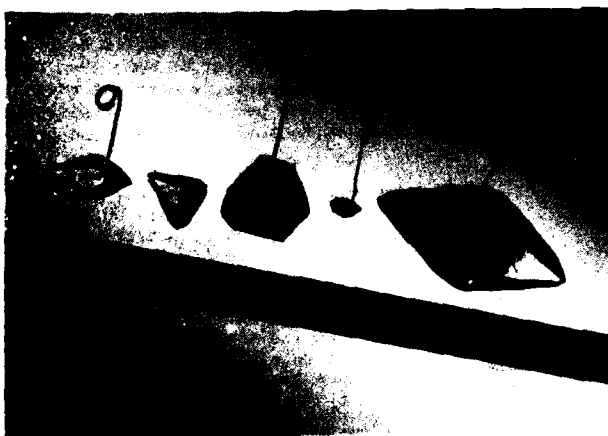


Figure 3 - Extreme left, two AlPO_4 single crystals grown on cut (0001) seeds. Center, large unused cut (0001) seed followed by small spurious nucleated seed. Extreme right view, large AlPO_4 crystal after five successive growth generations on a spurious nucleated seed similar to sample on left.

TEMPERATURE COMPENSATED CUTS OF BERLINITE AND β -EUCRYPTITE FOR SAW DEVICES

1Lt Robert M. O'Connell and Paul H. Carr
Deputy for Electronic Technology, Hanscom AFB, MA

Abstract

Calculations of the surface acoustic wave (SAW) properties of berlinite (AlPO_4) and β -eucryptite ($\beta\text{-LiAlSi}_2\text{O}_6$) show that both materials are temperature compensated along both singly rotated and doubly rotated orientations.

For berlinite, several orientations combine zero electromechanical power flow angles with more than four times the piezoelectric coupling of ST cut quartz. The best singly rotated orientation is a direct analog of the ST cut of quartz. This cut has, like ST quartz, a zero electromechanical power flow angle, but the distinct advantage of more than four times the piezoelectric coupling. Even more promising are the results obtained for two doubly rotated cuts which combine all the advantages of the singly rotated cut with the added feature of better diffraction properties than ST cut quartz.

For β -eucryptite, a singly rotated orientation has a SAW velocity of 3662 m/sec and almost twice the piezoelectric coupling of ST cut quartz, but the disadvantage of an electromechanical power flow angle of 18 degrees. On the other hand, a doubly rotated orientation has a zero electromechanical power flow angle, but only half the piezoelectric coupling of ST cut quartz.

The results for berlinite show that it represents an attractive alternative to quartz for use in broad-band, low insertion loss SAW devices. The relatively low piezoelectric coupling of β -eucryptite makes it unattractive for broad-band, low insertion loss applications, but its relatively large SAW velocity indicates that it may be useful in high frequency applications.

Introduction

The desire to develop improved broad-band, low insertion loss SAW devices with temperature independent performance characteristics has resulted in a search for substrate materials that are temperature compensated and have piezoelectric coupling greater than that of ST cut quartz. A major contribution to the search for such materials has been the development of a phenomenological model¹ which explains why known materials are temperature compensated. According to that model, temperature compensated materials possess either of the following anomalous properties: (1) a positive temperature coefficient of velocity or elastic constant or (2) a negative coefficient of thermal expansion. Quartz, for example, is temperature compensated² because the temperature coefficient of C_{66} , the elastic constant for shear propagation along the Z axis, is positive¹.

Because quartz is temperature compensated, recent attention has been focused on certain quartz derivatives. Two of these materials, berlinite (AlPO_4) and β -eucryptite ($\beta\text{-LiAlSi}_2\text{O}_6$), are particularly interesting because they each satisfy one of the criteria of the above-mentioned phenomenological model. Berlinite has, like quartz, a positive temperature coefficient for C_{66} , the elastic constant for shear propagation along the Z axis³. β -eucryptite, on the other hand, has a negative coefficient of thermal expansion in the direction of the hexagonal C-axis^{4,5}.

Recent calculations of the surface acoustic wave properties of berlinite and β -eucryptite have shown that both of these materials are temperature compensated

along both singly rotated and doubly rotated orientations. This paper will review the results of those calculations, consider their significance for SAW applications, and briefly compare them with results obtained with other materials being studied.

Berlinite

The procedure used to calculate surface acoustic wave properties has been described before^{6,7,8}. It requires, for the material being studied, experimental values for the elastic, piezoelectric, and dielectric constants, their respective temperature coefficients, the density, and the coefficients of thermal expansion. In the calculations reported here, the constants for berlinite were all taken from the data of Chang and Ba-sch³.

Initial results showed that berlinite indeed has temperature compensated cuts with more than four times the piezoelectric coupling of ST cut quartz⁶. However, none of those initially reported cuts had a zero electromechanical power flow angle, a desirable property characteristic of ST cut quartz. Subsequent calculations produced two singly rotated cuts and two doubly rotated cuts, all of which have zero electromechanical power flow angles⁷. The most promising of the singly rotated cuts is the X axis boule 80.4° cut, a direct analog of the ST cut of quartz. The two cuts are compared in Table I. Note that the piezoelectric coupling, $\Delta V/V^0$, of the X-axis boule 80.4° cut is more than four times as large as that of ST quartz, a distinct advantage. Other calculations have been made to investigate the behaviour of pseudo surface acoustic waves on berlinite⁹.

Table I also shows that the slope of the power flow angle, $\partial\phi/\partial\theta$, is larger for the X axis boule 80.4° cut of berlinite than it is for the ST cut of quartz. According to the theory of SAW diffraction¹⁰, this means that ST cut quartz has better diffraction properties than the X axis boule 80.4° cut of berlinite. The desire to find temperature compensated cuts of berlinite having better diffraction properties than the singly rotated 80.4° cut motivated consideration of doubly rotated cuts. In particular, the $\mu = 90.0$ plane of an orthogonal coordinate system having the three Euler angles λ , μ , and θ as its basis was carefully searched. The $\mu = 90.0$ plane was of particular interest because it contains four of the standard crystallographic cuts, including the X cut and Z axis cylinder, for which temperature compensated orientations were found earlier.

The results are shown in Figure 1. The dashed and solid curves represent, respectively, the loci of Euler angles for which the electromechanical power flow angle and the temperature coefficient of time delay are zero. As can be seen in the blown up portion of the Figure, the loci intersect in a total of twelve places throughout the plane. Because of crystal symmetry, however, only two of the points are independent, and those circled in Figure 1 are listed in Table I. Notice that while they have about the same piezoelectric coupling as the singly rotated 80.4° cut, the slopes of their power flow angles are smaller than those of either the 80.4° cut or the ST cut of quartz, giving them the added advantage of less diffraction spreading.

β -Eucryptite

To calculate the SAW properties of β -eucryptite, coefficients of thermal expansion were obtained from

the data of Hummel⁴ and Schulz⁵, and dielectric constants were obtained from the data of Bohm¹¹. Values for the elastic and piezoelectric constants and the density were obtained from the data of Barsch and Spear¹², which were measured on samples of β -eucryptite grown at Pennsylvania State University. Figure 2 shows the variation of SAW velocity, electromechanical power flow angle, piezoelectric coupling, and temperature coefficient of time delay for the X cut, for which a singly rotated temperature compensated orientation was found at 69° . Table I shows that although the piezoelectric coupling for this cut is almost twice as large as that of ST quartz, it has the disadvantage of an 18 degree electromechanical power flow angle.

As was done in the case of berlinite, doubly rotated cuts were considered also, and a temperature compensated cut having a zero electromechanical power flow angle was found in the $\lambda = 0.0^\circ$ plane, as shown in Figure 3. As can be seen in the Figure, the loci intersect in a total of four places throughout the plane. Again, because of crystal symmetry, only one of the points is independent, and it is listed in Table I where it can be seen that, unfortunately, the piezoelectric coupling of this doubly rotated cut is only about half as large as that of ST quartz. Perhaps the most attractive feature of this material is that it has the highest SAW velocity of all the temperature compensated materials listed in Table I, 3662 m/sec.

Other Temperature Compensated Materials

The sulfosalts are a class of materials of the form Tl_3BX_4 , where B can be V, Nb, or Ta, and X can be S or Se. Recent calculations have shown that at least two of these materials are temperature compensated with significantly larger piezoelectric coupling than berlinite^{13,14}. One particular cut of Tl_3VS_4 , for example, has four times the piezoelectric coupling of berlinite¹³. As shown in Table I, however, this cut has the disadvantage of a rather large electromechanical power flow angle, about -17 degrees. Another cut of the same material and one of Tl_3TaSe_4 , having zero electromechanical power flow angles, have also been found¹⁵. As the data in Table I shows, the piezoelectric coupling of these cuts is not as large as that of the first cut discussed, but it is still more than twice as large as that of berlinite. The table also shows that the SAW velocities of the sulfosalts are only about 1/3 as large as that of berlinite. This is a disadvantage for high frequency applications, but an advantage for long delay lines and low frequency SAW filters.

A composite material, consisting of a film of silicon dioxide on lithium tantalate, has also been shown to be temperature compensated¹⁶. This material has, as shown in Table I, a very small electromechanical power flow angle, a piezoelectric coupling of about .007, and a relatively large SAW velocity. The most attractive feature of the material is that its second order temperature coefficient of time delay is nearly an order of magnitude smaller than that of ST cut quartz. Despite these positive attributes, the composite has several drawbacks due to the SiO_2 film, including: (1) its thickness must be very accurately controlled, (2) it is very lossy at high frequencies, and (3) it is dispersive.

Conclusions

Several temperature compensated cuts of berlinite, having zero electromechanical power flow angles and more than four times the piezoelectric coupling of ST cut quartz, have been found. Even more encouraging are recent measurements of the piezoelectric strain constant d_{11} by X-ray methods which have shown that this quantity and, consequently, the piezoelectric coupling may be

20 to 30 percent larger than the previously determined values¹⁷. All these results indicate that berlinite appears to be a better substrate material than quartz, and as good quality supplies of the material become available, it will be attractive for use in broad-band, low insertion loss surface acoustic wave devices.

β -Eucryptite has also been shown to be temperature compensated, and this lends further credence to the phenomenological model¹ which helps predict which materials may be temperature compensated. The piezoelectric coupling of β -eucryptite is relatively poor, however, and it does not appear to be as attractive for broad-band, low insertion loss applications as berlinite; however, because of its relatively large SAW velocity, it may find use in high frequency applications.

Berlinite and β -eucryptite are shown on the state-of-the-art diagram in Figure 4, along with the other materials discussed above. Clearly, the search for high coupling temperature compensated materials has produced some attractive results, and it promises to produce more. No single one of these materials is perfect for every SAW application, but together they increase the variety of choices available to the design engineer and, most importantly, they remove the need to use lithium niobate with its associated ovens for broad-band, low insertion loss devices, for which ST quartz is not adequate.

References

1. R.E. Newnham, "Elastic Properties of Oxides and the Search for Temperature Compensated Materials," AFCRL Report No. TR-73-0220, Contract No. F19628-73-C-0108, 1973.
2. M.B. Schulz, B.J. Matsinger, and M.G. Holland, "Temperature Dependence of Surface Acoustic Wave Velocity on α -Quartz," J. Appl. Phys., Vol 41, pp 2755-2765, 1970.
3. Z.P. Chang and G.R. Barsch, "Elastic Constants and Thermal Expansion of Berlinite," IEEE Trans. on Sonics and Ultrasonics, Vol. SU-23, pp 127-135, 1976.
4. F.A. Hummel, "Thermal Expansion Properties of Some Synthetic Lithia Minerals," J. Am. Ceram. Soc. **34**, 235 (1951).
5. H. Schulz, "Thermal Expansion of Beta Eucryptite," J. Am. Ceram. Soc. **57**, 313 (1974).
6. P.H. Carr and R.M. O'Connell, "New Temperature Compensated Materials with High Piezoelectric Coupling," Proc. of the 30th Annual Symposium on Frequency Control, pp 129-131, June 1976.
7. R.M. O'Connell and P.H. Carr, "High Piezoelectric Coupling, Temperature Compensated Cuts of Berlinite, $AlPO_4$, for SAW Applications," IEEE Trans. on Sonics and Ultrasonics (to be published).
8. A.J. Slobodnik, Jr., "The Temperature Coefficients of Acoustic Surface Wave Velocity and Delay on Lithium Niobate, Lithium Tantalate, Quartz, and Tellurium Dioxide," AFCRL Physical Sciences Research Paper No. 477, AFCRL-72-0082, 1971.
9. A. Jhunjhunwala, J.F. Vetolino, and J.C. Field, "Berlinite, A Temperature Compensated Material for Surface Acoustic Wave Applications," Proc. of the 1976 Ultrasonics Symposium, pp. 523-527, Sep 1976.
10. T.L. Szabo and A.J. Slobodnik, Jr., "The Effect of Diffraction on the Design of Acoustic Surface Wave Devices," IEEE Trans. on Sonics & Ultrasonics,

Vol. SU-20, pp 240-251, 1973.

11. H. Bohm, "Dielectric Properties of β -Eucryptite," Phys.Stat.Sol., (a) **30**, 531 (1975).
12. G.R. Barsch and K.E. Spear, "Temperature Compensated Piezoelectric Materials," AFCRL Report No. TR-75-0609, Contract No. F19628-75-C-0085, 1975.
13. R.W. Weinert and T.J. Isaacs, "New Piezoelectric Materials Which Exhibit Temperature Stability for Surface Waves," Proc. of the 29th Annual Symposium on Frequency Control, pp. 139-142, 1975.
14. T.J. Isaacs and R.W. Weinert, "Crystal Growth and Properties of Ti_3Bx_4 Crystals for Acoustic Surface-Wave and Bulk Acoustic Devices," Journal of Electronics Materials, Vol 5, No. 1, pp 13-22, 1976.
15. A. Jhunjhunwala, J.F. Vetelino, and J.C. Field, "Temperature Compensated Cuts with Zero Power Flow in Ti_3VS_4 and Ti_3TaSe_4 ," Electronics Letters, Vol. 12, No. 25, pp 683-684, 1976.
16. T.E. Parker and M.B. Shulz, "Stability of SAW Controlled Oscillators," Proc. of the 1975 Ultrasonics Symposium, pp. 261-263, 1975.
17. G.R. Barsch and K.E. Spear, "Temperature Compensated Piezoelectric Materials," RADC-TR-77, Final Report, Contract No. F19628-75-C-0085 (May 1977).

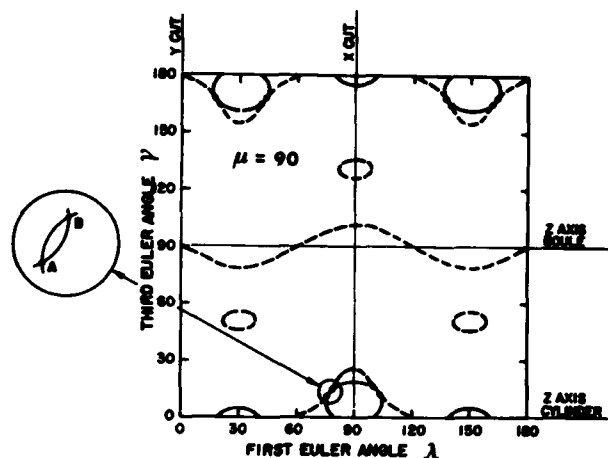
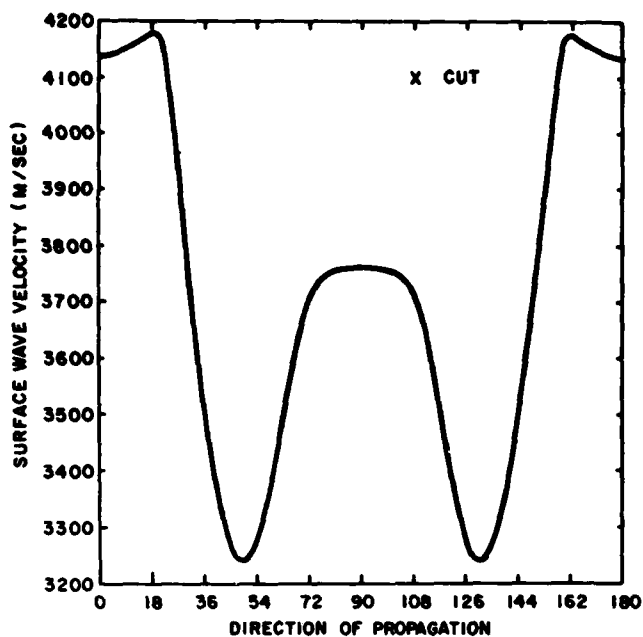


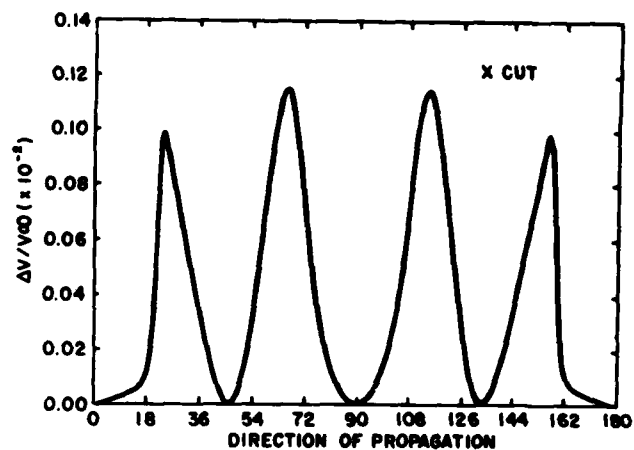
Figure 1 - Loci of Euler angles having zero electromechanical power flow angle (dashed lines) and zero temperature coefficient of time delay (solid lines) in the $\mu = 90.0$ plane of berlinite.

MATERIAL	ORIENTATION	EULER ANGLES λ μ θ			POWER FLOW ANGLE ϕ (DEG)	SLOPE OF POWER FLOW ANGLE ($d\phi/d\theta$)	$\Delta v/v_{90}$ ($\times 10^{-2}$)	SAW VELOCITY (m/sec)
QUARTZ (SiO_2)	ST CUT	0	132.75	0	0.0	0.378	0.058	3158
BERLINITE ($AlPO_4$)	X AXIS BOULE 80.4°	0	80.4	0	0.0	0.901	0.245	2751
	DOUBLY ROTATED, A	76.8	90	11.5	0.0	0.372	0.250	2756
	DOUBLY ROTATED, B	79.7	90	15.5	0.0	0.221	0.247	2758
β -EUCRYPTITE (β - $LiAlSiO_4$)	X CUT 69°	90	90	69	18	--	0.100	3662
	DOUBLY ROTATED	0	57	62	0.0	0.32	0.035	3258
Ti_3VS_4	(110) CUT 70°	-45	90	70	-17	--	1.0	900
	(110) CYLINDER 24°	45	24	90	0.0	--	0.617	1010
Ti_3TaSe_4	(110) CYLINDER 54°	45	54	90	0.0	--	0.508	879
$SiO_2/LiTaO_3$	Y CUT, Z PROP	0	90	90	0.0	--	0.7	3455

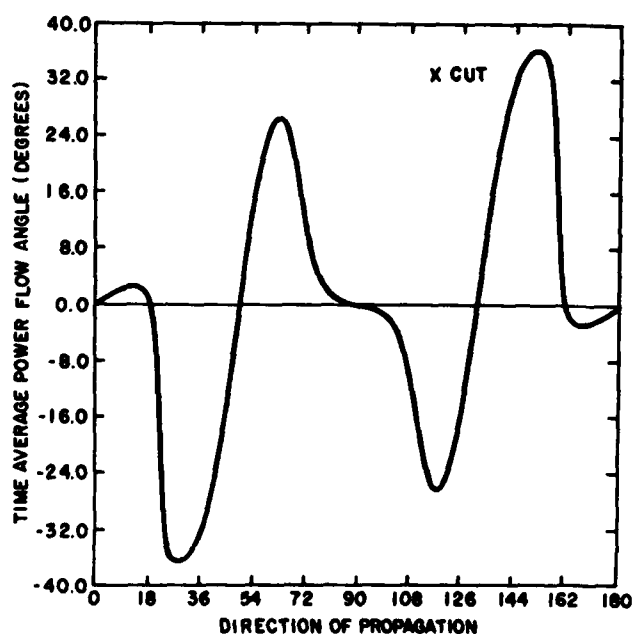
TABLE 1. TEMPERATURE COMPENSATED CUTS OF VARIOUS MATERIALS



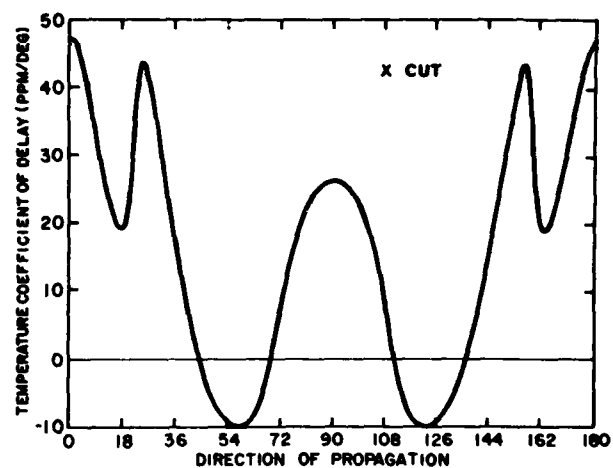
(a) SAW velocity



(c) Piezoelectric coupling



(b) Electromechanical power flow angle



(d) Temperature coefficient of time delay for X-cut β -eucryptite

Figure 2 - The variation of (a) SAW velocity, (b) electromechanical power flow angle, (c) piezoelectric coupling, and (3) temperature coefficient of time delay for X-cut β -eucryptite.

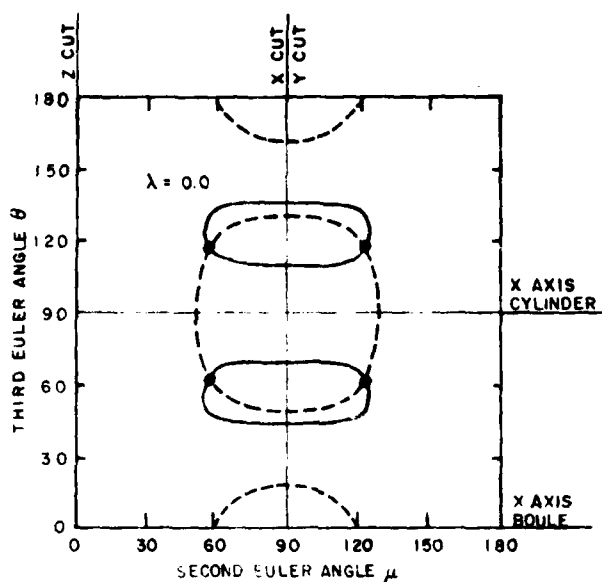


Figure 3 - Loci of Euler angles having zero electromechanical power flow angle (dashed lines) and zero temperature coefficient of time delay (solid lines) in the $\lambda = 0.0$ plane of β -eucryptite.

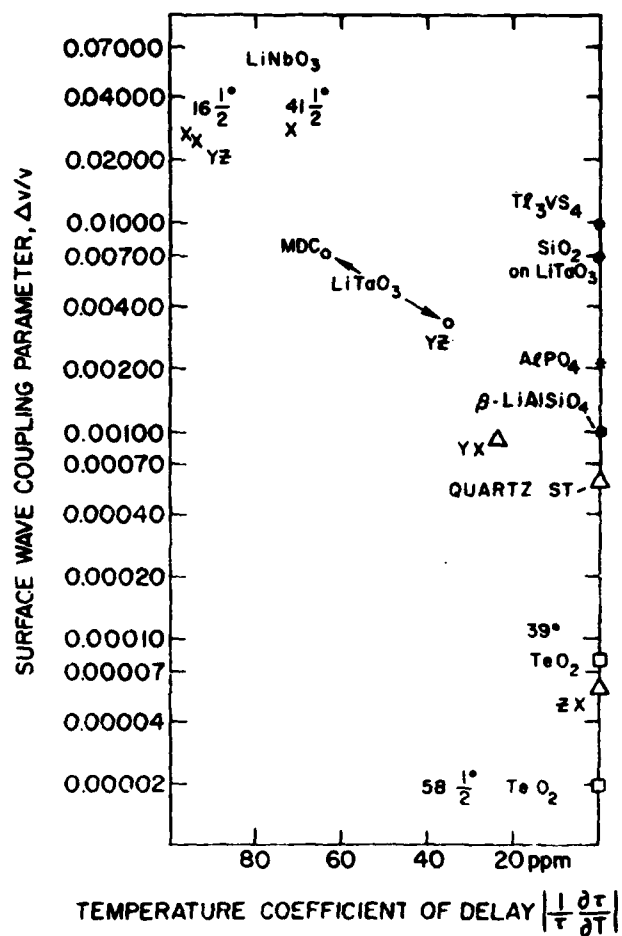


Figure 4 - Temperature coefficient of time delay versus piezoelectric coupling for various SAW materials.

CCD RECURSIVE FILTER FOR MTI APPLICATIONS*

William L. Eversole, Walter H. Bailey, Patrick L. Ham
Texas Instruments Incorporated
Dallas, Texas

Summary

The evolution of charge coupled devices (CCDs) has provided the system designer with a new arsenal of design tools. The simplest CCD building block, the CCD delay line, offers the designer an electronically controllable solid state analog shift register. This paper discusses the uses of CCD delay lines for recursive filtering and the application of the recursive architecture for MTI filtering for radars. Some of the limitations of CCDs are discussed along with methods of overcoming some of these limitations.

Introduction

Radar video processing is an area where significant improvements in radar performance may be achieved by exploiting recent technology advances. The general radar video processing problem is outlined in Figure 1 which shows returns from the k th range bin which include an aircraft, wind driven precipitation, and ground returns. The spectrum of the returns for the k th range bin between DC and the PRF is also shown in this figure. These typical spectral characteristics indicate the desirability of some form of target classification based upon spectral characteristics rather than upon amplitude alone.

One approach to target classification is the Moving Target Indicator (MTI). The purpose of the MTI is to reject signals from fixed unwanted targets, such as buildings, trees, etc., and retain for detection the signals from moving targets such as aircraft. The MTI is a sampled data filter and can be implemented with CCD delay lines. The degree of sophistication of MTI filters varies from simple single delay feed-forward cancellers to multi-pole recursive filters.

CCD Recursive Filter

In charge coupled devices (CCDs) the analog input signal is sampled and stored as discrete analog packets of charge and hence ideally suited to a number of sampled-data signal processing functions. Recursive filtering architecture is advantageous over transversal filtering when a sharp transition bandwidth is required in the frequency characteristic or where a large amount of time multiplexed data is to be filtered.

A general N th order recursive filter has the transfer function:¹

$$H(z) = \frac{\sum_{i=0}^N a_i z^{-i}}{\sum_{i=0}^N b_i z^{-i}} \quad (1)$$

CCDs can be used to implement the delay term, z^{-1} , in equation 1. The delay of an N -stage CCD delay line clocked at f_c is

$$T_d = N/f_c \quad (2)$$

Since a CCD delay line samples the input signal at the clock rate, the sampling rate, f_s , is equal to the clock rate, f_c , and this determines the Nyquist bandwidth. The stability criteria for CCD recursive filters is the same as any sampled data filter in that all poles must lie inside the unit circle of the Z -plane.

CCDs have performance limitations, two of which are imperfect charge transfer efficiency and leakage current. The effect of charge transfer efficiency (CTE) on the transfer function for a N -stage CCD delay line is to multiply the ideal delay, z^{-N} , by a dispersion factor:²

$$D(z) = \left[\frac{1-\epsilon}{1-\epsilon z^{-1}} \right]^N \quad (3)$$

where ϵ is the charge loss per stage. Imperfect CTE can change the filter characteristics.³

Thermal leakage limits the maximum time delay which can be achieved by CCDs.⁴ The leakage current adds to the signal charge and the size of the signal must be reduced to avoid saturation of the CCD potential wells. This reduces the dynamic range of the CCD. Since leakage current increases with temperature, applications where long time delays are needed may require temperature control.

MTI Applications

One of the most important applications for CCD recursive filters is the delay canceller in MTI radar.^{5,6} In radar systems, the returns from the stationary or slowly moving targets are frequently larger than the returns from moving targets, therefore a moving target indicator, MTI, is needed to either enhance the signals from the moving targets or suppress stationary target returns. Since clutter or stationary targets are characterized by zero doppler shift, their returns are periodic spectra at DC and multiples of the PRF, while moving targets have returns shifted by their doppler frequencies. Figure 1 shows the relation between clutter and a typical moving target. Because this filtering must be performed on up to a few thousand range bins simultaneously, recursive filtering is attractive. This recursive filter can be realized using digital approaches but the cost and complexity of analog-to-digital conversion may be prohibitive. Acoustic delay lines can be used to implement the delay stages in the recursive filter transfer function but controlling the delay time accurately is a problem. Charge coupled device (CCD) signal processing eliminates these problems of the digital and acoustic delay line approaches to recursive filters. CCDs take advantage of the precise timing of digital processing techniques without the requirement for analog-to-digital conversion. The stability of the CCD filter characteristics is determined by precise delay times which are achieved by synchronously clocking the CCD from signals derived from the radar's master clock.

* This work is supported by ECOM under contract DAAB07-76-C-0912.

The radar characteristics impose several requirements on the MTI filter in addition to rejection of clutter while passing moving target returns without distortion. The bandwidth and sampling frequency of the MTI filter are determined from the radar pulse width, τ , which also determines the range resolution.

$$\text{Bandwidth} \quad BW \geq \frac{1}{2\tau} \quad (4)$$

$$\text{Sampling Frequency} \quad f_s \geq \frac{1}{\tau} \quad (5)$$

$$\text{Range Resolution} \quad \delta_R = \frac{c\tau}{2} \quad (6)$$

Since range resolution is determined by the radar pulse width, τ , high range resolution dictates small radar pulse width, which increases the sampling frequency, f_s . The sampling frequency determines the number of delay stages, N , needed in the CTD.

$$N = f_s T_d \quad (7)$$

where T_d is the total delay of the CCD ($T_d = \text{PRI}$). From this equation, it is seen that increased sampling frequency requires longer delay lines for a given PRF. If long CCD delay lines are used, the dispersion caused by imperfect CTE of the analog signal could yield undesired responses due to uncanceled residues from stationary targets. Rather than using long delay lines, the radar return can be demultiplexed into several parallel filters.⁵ When using a multiplexed approach, the number of delay stages for each delay line is reduced by $1/M$ where M is the number of parallel filters. A block diagram of a multiplexed system using fourth order filters is shown in Figure 2.

Experimental Results

A four pole, five pulse MTI canceller has been constructed using 150 bit, 2 phase N-channel CCDs. The MTI filter is a cascade of two second order recursive sections and has transfer function given by:

$$H(z) = \frac{(1-z^{-1})^4}{(1-\alpha_{11}z^{-1} - \alpha_{21}z^{-2})(1-\alpha_{12}z^{-1} - \alpha_{22}z^{-2})} \quad (8)$$

A block diagram of the filter is shown in Figure 3. The feedback coefficients were calculated to provide the filter with a Butterworth response. The feed-forward and feedback coefficients are implemented with variable gain operational amplifiers. The filter also has electronically selectable cutoff frequencies implemented with analog switches and resistors. A photograph of the frequency response is shown in Figure 4. The responses at zero frequency are due to the zero frequency marker in the spectrum analyzer. The response of the MTI filter to two different frequencies is shown in Figure 5. The top photograph of Figure 5 shows the spectrum of the input to the MTI filter. There are two frequencies present, one in the passband of the filter, the other one at the PRF. The bottom photograph shows the output of the MTI filters with the frequency in the passband still present while the frequency at the PRF has been cancelled.

Conclusions

The application of CCD delay lines to recursive filtering has been described. A four pole MTI filter with electronically selectable cutoff frequencies and a cancellation ratio of 40 dB has been discussed.

Because of the commercial availability of CCD delay lines, many more recursive filter applications can be realized. Also, as more work is directed toward integrating the peripheral circuitry the cost, power and weight of systems will go down while the performance will increase.

References

1. L.R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*, Prentice-Hall, 1975.
2. D.D. Buss, A.F. Tasch, Jr., and J.B. Barton, "Applications of CCDs to Analog Signal Processing," *Solid-State Devices and Circuits, Vol. II*, Wiley-Interscience. (To be published).
3. J.E. Bounden, R. Eames, and J.B.G. Roberts, "MTI Filtering for Radar with Charge Transfer Devices," *Proc. CCD Technology and Applications Conf.*, pp 206-213, Edinburgh, September 1974.
4. A.F. Tasch, Jr., R.W. Broderson, D.D. Buss, and R.T. Bate, "Dark-Current and Storage-Time Considerations in Charge-Coupled Devices", *Proc. CCD Applications Conf.*, pp 179-187, San Diego, Sept. 1973.
5. W.J. Butler, W.E. Engeler, H.S. Goldberg, C.M. Puckette, IV., and H. Lobenstein, "Charge-Transfer Analog Memories for Radar and ECM Systems", *IEEE S. Solid-State Circuits*, SC-11, pp 93-100, February 1976.
6. J.E. Bounden and M.J. Tomlinson, "CCD Chebyshev Filter for Radar MTI Applications", *Electronics Letters*, 10, pp 89-90, April 1974.

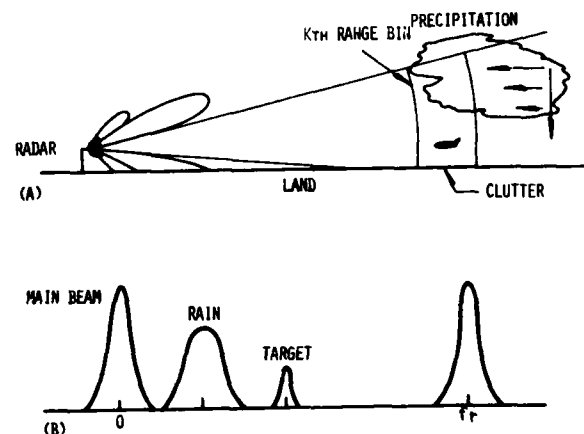


Figure 1

Relationship Between (A) Radar Operating Environment and (B) Spectral Characteristics

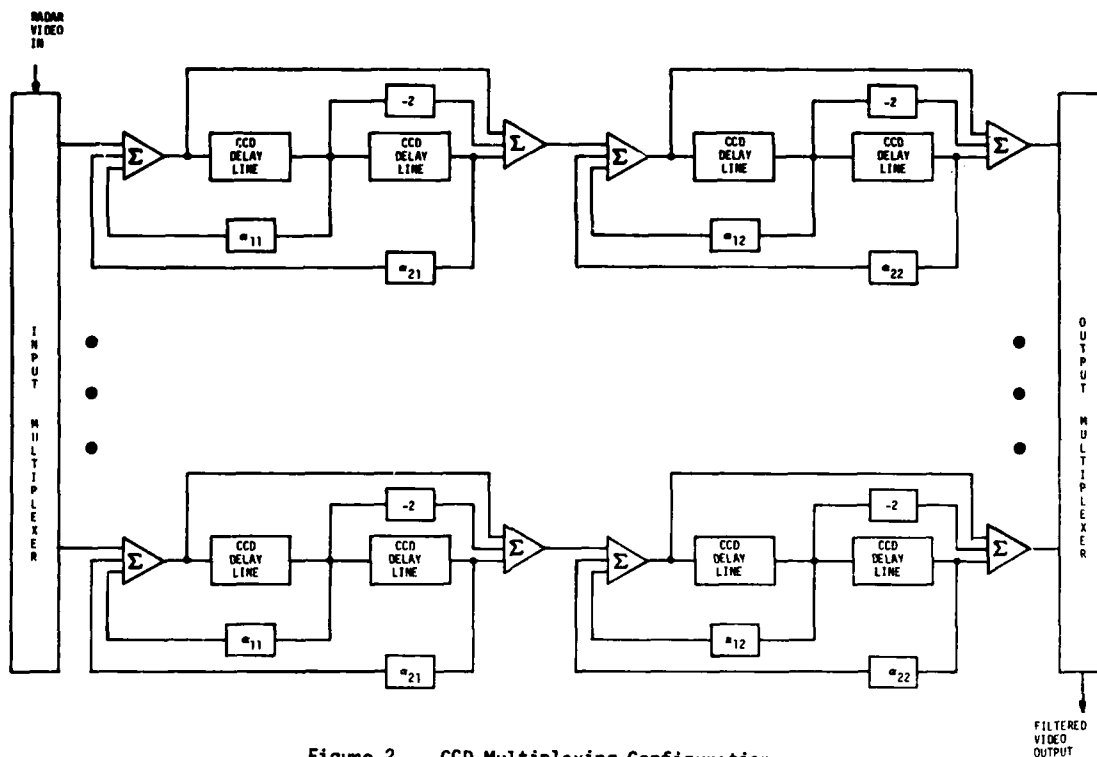
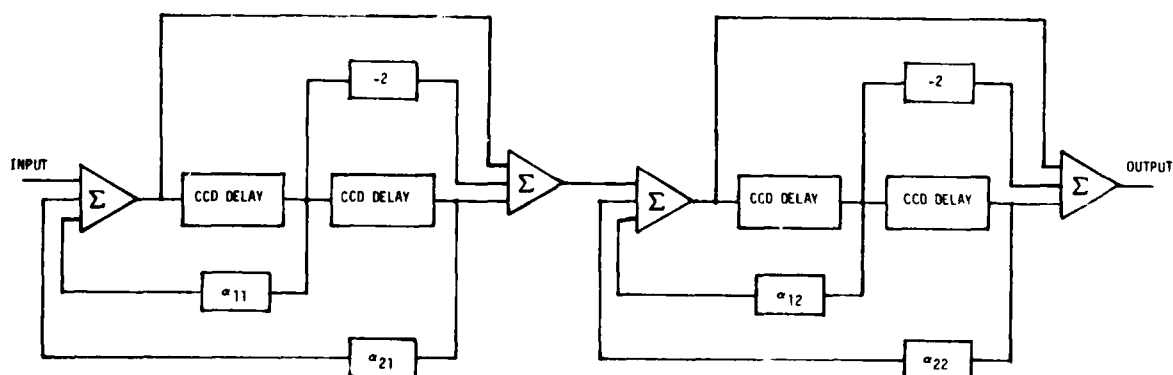


Figure 2 CCD Multiplexing Configuration



$$H(z) = \frac{(1 - z^{-1})^4}{(1 - a_{11} z^{-1} - a_{21} z^{-2})(1 - a_{12} z^{-1} - a_{22} z^{-2})}$$

Figure 3 Cascade Realization of a Fourth-Order System

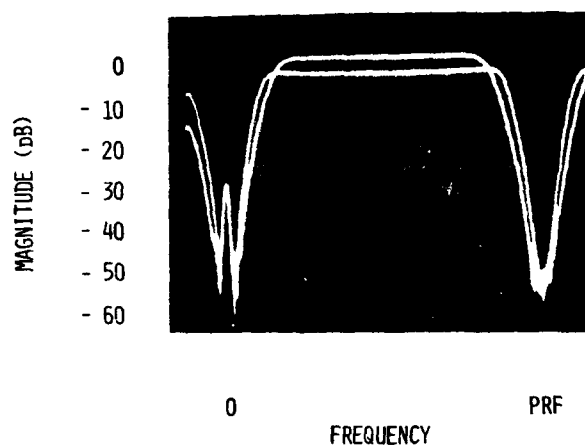


Figure 4 MTI Filter Frequency Responses

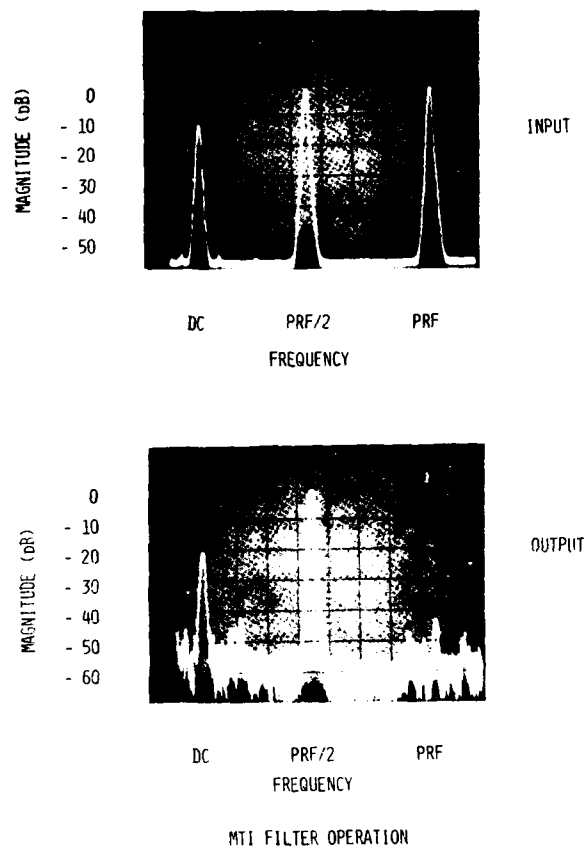


Figure 5 MTI Filter Operation

UNWANTED MODES IN MONOLITHIC CRYSTAL FILTERS

G. T. Pearman and R. C. Rennick
Bell Telephone Laboratories, Incorporated
Allentown, Pennsylvania 18103

Summary

In designing an AT-cut quartz thickness-shear/twist monolithic crystal filter, three frequency ranges must be considered: the pass-band, the transition-band, and the stop-band. This paper is not concerned with the pass-band which is controlled primarily by the tolerance placed on resonator frequencies and inter-resonator couplings. Instead, the design considerations for realization of the transition-band and stop-band performance will be discussed.

The order of the filter (number of resonators) will determine the ultimate performance in the transition-band and stop-band if the design does not employ transmission zeroes. But if this performance is to be realized, the designer must deal with the unwanted modes of vibration in the monolithic crystal filters.

Unwanted modes in a monolithic crystal filter can be classified into two general groups: the trapped inharmonic modes of thickness-shear/twist and the overtones of what are called plate modes in this paper. The first group of modes occur at frequencies between the plated resonator frequencies and the unplated plate frequencies and are affected to a great degree by the resonator dimensions. The second group consists of flexure, face-shear, and extensional modes which depend primarily on the plate lateral dimensions rather than on the resonator shapes. These modes occur at many discrete frequencies and, although they are usually weaker than the trapped inharmonic modes, it is shown that they may seriously degrade the filter performance.

A method of designing a resonator array for minimizing the response level of the inharmonic modes will be described. This method relies on predictions of individual resonator inharmonic responses as well as those predicted for the total structure. These predictions will be compared with data obtained from actual models made using several different eight resonator arrays.

The effect of plate modes, especially in the transition-band, will be shown for these same models. The identification of the plate modes can be difficult in practice due to the large number and relative weakness of the responses although estimates can be made using standard references. Methods of identifying these modes will be illustrated. Experimental results are found to be in good agreement with calculations by T. R. Meeker. Once the modes have been identified, it will be shown how the designer can control the effect of these modes in the frequency range of interest.

Introduction

Monolithic crystal filters--indeed all filters-- have three frequency bands of interest: the pass-band, the stop-band, and the transition-band. This paper will not be concerned with the pass-band but instead will deal with the effects of unwanted modes on the stop- and transition-bands of monolithic crystal filters. Methods of minimizing these effects will be presented.

An eight resonator, AT-cut monolithic crystal filter of the type pictured in Figure 1 will be used as an example throughout. This filter has a center frequency of about 8.150 MHz and a nominal 3 dB bandwidth of 3.25 KHz. The z -axis is used as the coupling direction (the long plate dimension) with the x -axis (diagonal axis) being along the plate width. The nominal inductance is 30 mH although, as can be seen in the photograph, some resonators are designed to vary by as much as $\pm 10\%$.

For purposes of discussion, the unwanted modes are classified into two groups: first, the trapped inharmonic overtones of thickness-shear/twist and, second, what are called plate modes. These latter modes consist of the overtones of flexural, face-shear, and extensional vibrations. The insertion-loss versus frequency scan of a filter shown in Figure 2 illustrates the effects of these modes. The sharp resonances occurring immediately above the pass-band are the inharmonic modes. The discontinuity seen in the lower transition-band is due to a plate mode.

Shown in Figure 3 is the insertion-loss versus frequency performance of a monolithic crystal filter having eight identical resonators. The photographs are X-ray topographs of a portion of the filter. The dark regions are those areas where the crystal lattice is vibrating. These modes can be grouped according to the number of half-wavelengths in the width and length directions under the individual resonators. For example, the strong group of modes at 8.290 MHz are the (3, 1) modes--three half-wavelengths in the width direction and one half-wavelength along the length. Modes such as those around 8.408 MHz are a type of untrapped inharmonic mode. These modes are suppressed when non-identical resonators are used in the resonator array.

Since the inharmonic modes occur at frequencies which are partially determined by the electrode dimensions, the response levels can be minimized by properly designing the array. As graphically demonstrated by this particular filter design, the use of identical resonator shapes is not the proper design. In fact, the use of varied

resonator shapes having no inharmonic responses occurring at common frequencies proves to be the simplest method of achieving good inharmonic suppression. Briefly, the technique is as follows. The inharmonic mode frequencies of a variety of individual resonator shapes are calculated from a theoretical model which neglects any effects of flexure. This is done on a resonator-by-resonator basis. From these calculations, resonators are chosen for the filter array avoiding any mode frequency coincidences. Using the same theoretical model, the inter-resonator spacings necessary to achieve the correct couplings are then determined. Finally, the inharmonic mode spectrum of the total array is calculated.

The technique is illustrated with the following example. In Figure 4, "tic marks" indicate the frequencies of the three types of trapped inharmonic modes for the eight individual resonators of the filter seen in Figure 1. The numbers under the marks represent an order of the electrodes in this array. Resonators 1 and 6 are identical in this example and so have the same mode spectra. Several frequency coincidences between modes can be observed.

After determining the required inter-resonator spacings, the mode spectrum of the resultant array is calculated. The frequency spectrum of an actual filter of this design is shown above the calculated spectrum. The coincidence in frequency of the two modes at 117 KHz above center frequency results in a very strong mode--that mode where the two dashed lines meet. Figure 6 shows the insertion-loss versus frequency scan of a filter of this design. There is good agreement between the calculated mode spectrum of Figure 5 and the actual response of Figure 6 which also clearly shows the predicted strong mode.

Figures 5 and 6 indicate that the region from approximately 60 to 80 KHz above the pass-band is free of unwanted modes. In order to improve the inharmonic suppression, two resonators (#2 and #6) having responses at 117 KHz above center frequency are replaced with resonators having responses in the open region. Figure 7 shows a scan for a filter with the two resonators changed. The two modes which have been added to the formerly open region are easily seen when comparing Figures 6 and 7. Since this design avoids mode coincidences, the out-of-band suppression level has improved considerably.

That is, in brief, a way in which inharmonic modes can be suppressed. The method does not depend upon calculation of mode strengths which, in the past, have proved to be somewhat unreliable. It does provide a relatively straightforward way which has been successful in achieving good inharmonic mode suppression.

Plate Modes

The second class of modes to be considered are what are often called plate modes. Recall that these are the overtones

of flexural, face-shear, and extensional modes. They occur at nearly harmonically related frequencies both above and below the pass-band. When one or more of these modes fall near the pass-band of a filter, distortion in the transition-band can result as shown in Figure 8. Rather than rising monotonically as expected from the filter design, a "rolling-off" is seen in the lower transition-band of the filter. This example is a rather extreme one. In many cases, the presence of an interfering mode is not so evident--the roll-off is less noticeable although still present.

Whereas the inharmonic modes have frequencies which are strongly affected by the resonator shapes, the plate modes are relatively insensitive to these shapes. Instead, their frequencies are determined primarily by the lateral dimensions of the plate. Hence the designation, "plate modes"--and the key to the solution of the problem.

The existence of these modes is illustrated in Figure 9. The insertion-loss versus frequency characteristics of this same filter over a three plus MHz range are shown. A great many modes are found, the strongest of which have a 74 KHz spacing. Note that these modes are very weak, being suppressed around 70 dB relative to the in-band loss. Nonetheless, when they happen to fall close to the pass-band, they cause transition-band distortion.

That the cause of transition-band distortion is an interfering plate mode becomes even clearer in Figure 10. In this insertion-loss versus frequency scan the frequencies of a series of modes below the pass-band are marked. If 74 KHz is added to the 8.076 MHz of the mode closest to the pass-band, the frequency of the next mode would be expected to be around 8.150 MHz. This is exactly where the transition-band rolls off.

A literature search reveals that, for a plate the size of that used in this monolithic crystal filter (34.4 mm in length), length dependent face-shear modes should have a spacing of approximately 74 KHz.⁽¹⁾ The face-shear mode causing the problem seen in this filter at 8.150 MHz is the 110th overtone!

Several experiments were performed to verify the identification of these unwanted plate modes. Application of damping material to the ends of the plate should greatly attenuate the length dependent face-shear mode. Figure 11 shows the effect of applying a thin line of epoxy across the ends of the plate. The strong modes (74 KHz) are greatly attenuated. Most of the modes which still exist are width dependent. The family of flexural modes were identified, some of which are noted with small arrows. These modes have a spacing of around 144 KHz, again in good agreement with that predicted from available literature for a plate of this width (11.2 mm). When epoxy is applied to the width edges of the plate, the responses marked by the small arrows are eliminated.

Another experiment which can be done to verify the classification of these modes (especially the weaker flexural modes) is to design and make an x-coupled filter. To do this, the plate and array are rotated 90° making the long dimension (the coupling direction) the x or diagonal axis. Then, of course, the z'-axis becomes the width direction. Since the primary frequency controlling axis (x) for the flexural mode is now about three times as long as in the z'-coupled case, the flexural mode spacings should be one-third that found in the z'-coupled case (or 48 KHz). Figure 12 shows the performance of one such filter. The two modes below the pass-band and certain of those above are flexural modes with a spacing of about 47 KHz. In this case, the flexural modes are considerably stronger than observed in the z'-coupled filters while the face-shear modes are suppressed. This change in suppression levels is probably due to two causes. First, the designs are symmetric along the width (and not along the length) and, second, the plates are damped by the mounting structure on the width edges as shown in Figure 1. Both facts may result in suppression of the width dependent modes whether they be flexure or face-shear. At any rate, the spacings agree well with those expected.

T. R. Meeker has recently described a method of calculating these mode frequencies⁽²⁾ using equations derived from the work of Mindlin and Spencer⁽³⁾ and Tiersten.⁽⁴⁾ The results of this study agree quite well with Meeker's predictions also.

Given the fact that an interfering mode is causing a problem at a particular frequency, the filter center frequency can be adjusted by either addition or removal of mass from the electrodes resulting in an improved filter at a different frequency. Figure 13 shows the result of ion milling one filter in two 5 KHz steps. As the center frequency is raised by the milling operation, the transition-band improves. The frequency of the interfering mode is unchanged as expected.

Conversely, if the roll-off in the transition-band is due to plate modes, one should be able to change the plate dimensions and improve the performance of a filter at a given frequency. Figure 14 shows the effect on the transition-band of shortening the plate length in two 0.025 mm steps. As we move the interfering mode (in this case, a length dependent face-shear mode), the lower transition-band performance improves.

In practice, of course, these modes should be placed outside the region of interest by correctly dimensioning the plate. Although it is a relatively simple matter to determine the frequency spacing between overtones of plate modes, it may be difficult to predict the precise frequency at which these high overtones will fall. These precise frequencies depend on the effective acoustic dimensions of the plate which may differ from the mechanical dimensions due to the plate edge finish or the mounting structure, for example. For the situation

where filters are required over a frequency range larger than the spacings between modes, two or more plate sizes may be needed to ensure freedom from the interfering plate modes.

Closing Remarks

The investigation has concentrated on the phenomena of interfering plate modes only for rectangular plates. For circular plates, the mounting structure is usually attached to the plate at a 45° angle to the x-axis. The combination of mechanical damping through the mounting wires and the circular plate geometry probably provides adequate suppression of these modes. Plate modes are usually not a problem.

For the inharmonic modes, the comments herein apply to both circular and rectangular plates since these modes are affected largely by the electrode dimensions--not the plate geometry. The methods of achieving inharmonic mode suppression would be identical.

References

1. R. A. Heising, "Quartz Crystals for Electrical Circuits", pp. 205-248, D. Van Nostrand Company, Incorporated (1946).
2. T. R. Meeker, "X1 and X3 Flexure, Face-Shear, Extension, Thickness Shear, and Thickness Twist Modes in Rectangular Rotated Y-Cut Quartz Plates", these proceedings.
3. R. D. Mindlin and W. J. Spencer, "Anharmonic Thickness-Twist Overtones of Thickness-Shear and Flexural Vibrations of Rectangular AT-Cut Quartz Plates", J. Acoust. Soc. Am. 42, pp. 1268-1277 (1967).
4. H. F. Tiersten, "Linear Piezoelectric Plate Vibrations", Plenum Press (1969).

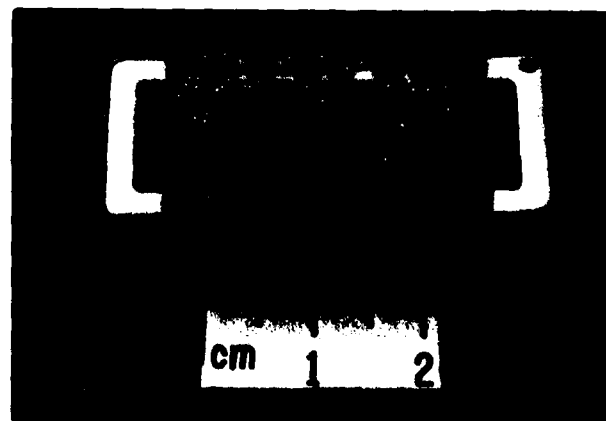


Figure 1 - 8 Resonator Monolithic Crystal Filter.

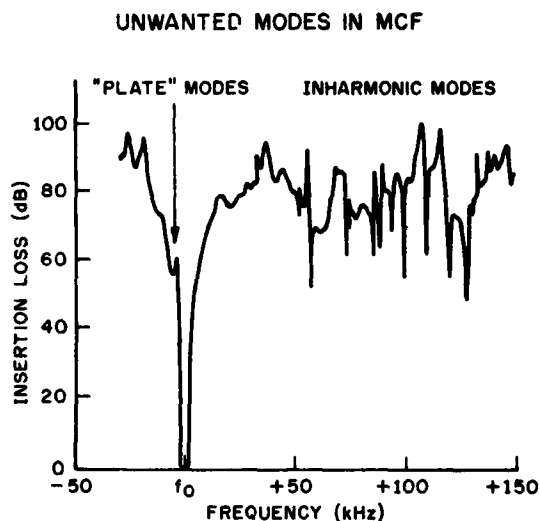


Figure 2
Insertion-Loss Versus Frequency Scan of
8 Resonator Monolithic Crystal Filter.

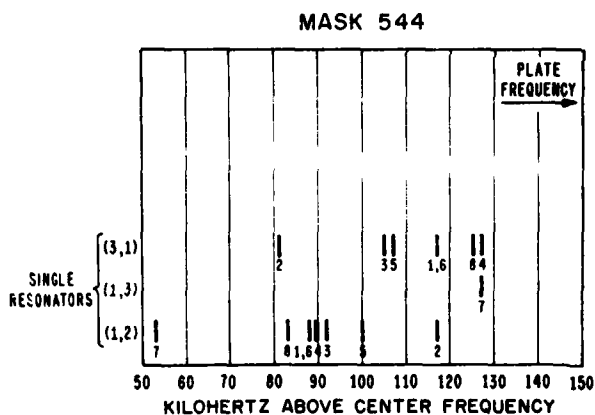


Figure 4
Resonator Inharmonic Mode Spectra for 8
Resonator Monolithic Crystal Filter.

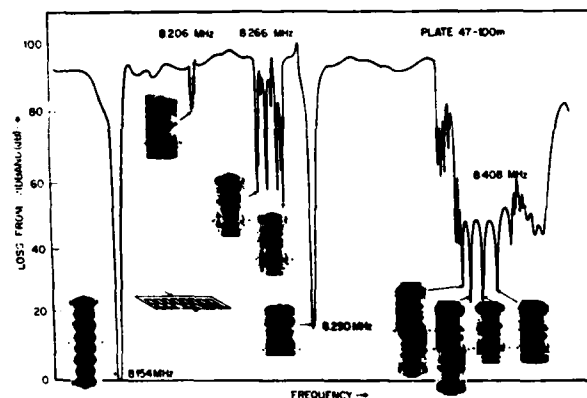
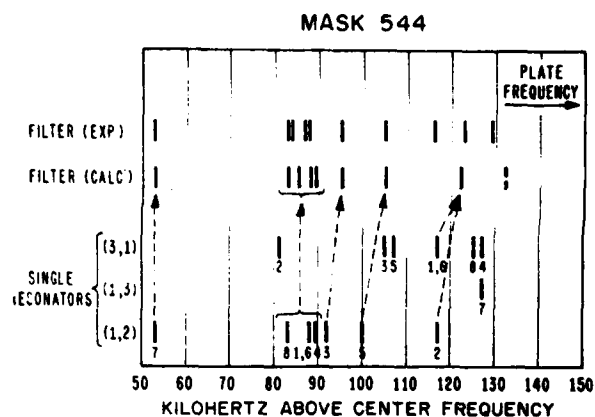


Figure 3
Insertion-Loss Versus Frequency Scan of 8
Identical Resonator Monolithic Crystal Filter
(X-ray Topographs of Mode Patterns Included).



Same as Figure 4 With Comparison of Calculated
and Observed 8 Resonator Monolithic Crystal
Filter Inharmonic Spectra.

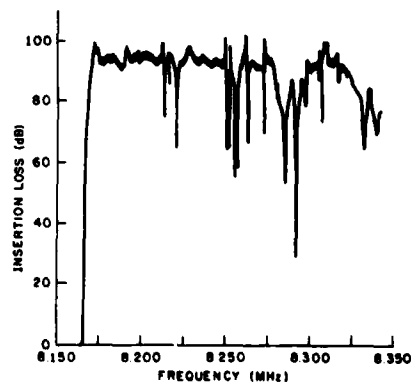


Figure 6
Insertion-Loss Versus Frequency Performance of
the 8 Resonator Monolithic Crystal Filter Shown
in Figure 1.

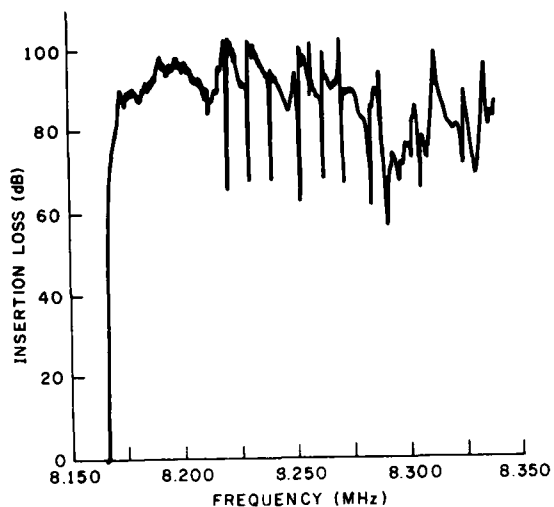


Figure 7
Insertion-Loss Versus Frequency Performance of
Improved Design Monolithic Crystal Filter.

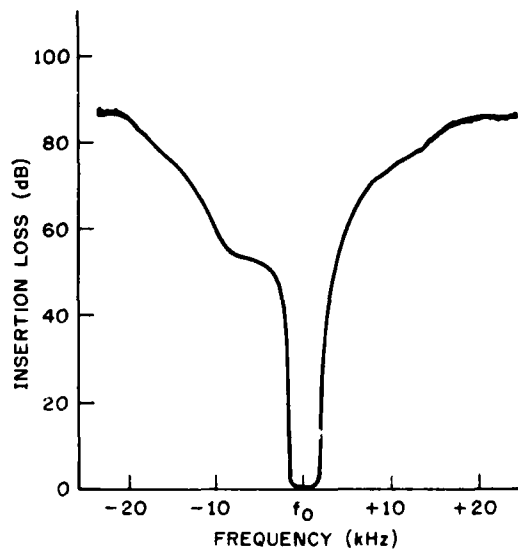


Figure 8
Pass-Band/Transition-Band Response of 8
Resonator Monolithic Crystal Filter.

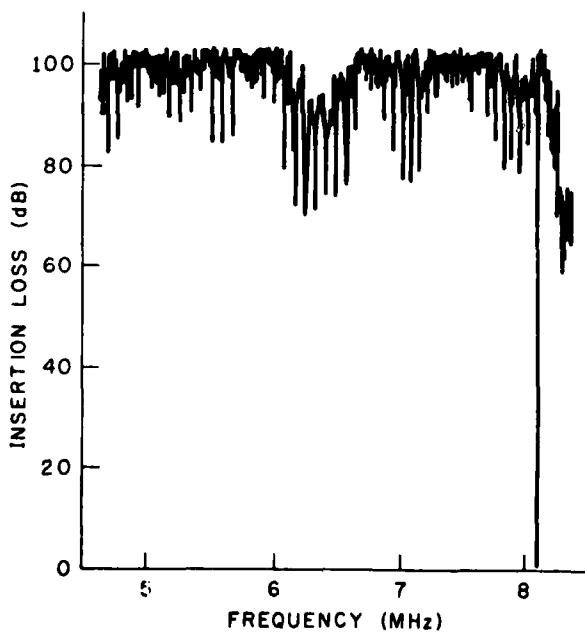


Figure 9
Plate Mode Spectrum Observed Over 3 MHz Range
for an 8 Resonator Monolithic Crystal Filter.

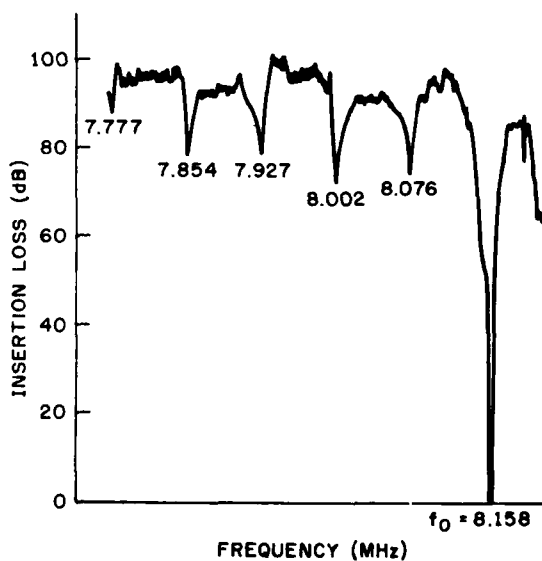


Figure 10
Detailed Plate Mode Spectrum of 8 MHz
Monolithic Crystal Filter.

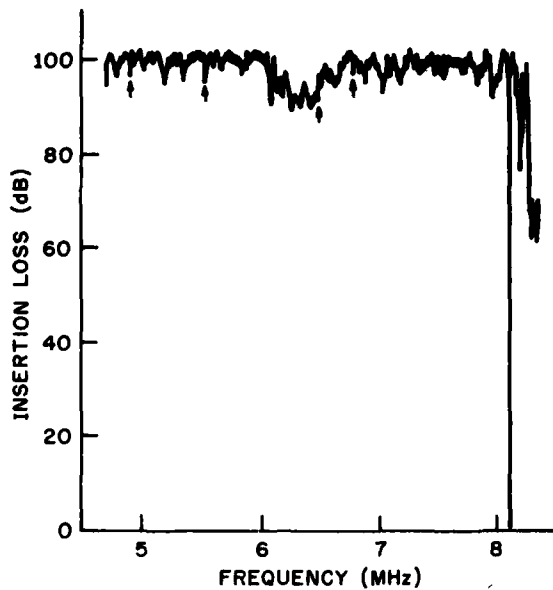


Figure 11
Plate Mode Spectrum Observed Over 3 MHz Range
for Monolithic Crystal Filter with Damping
Material on Plate Ends.

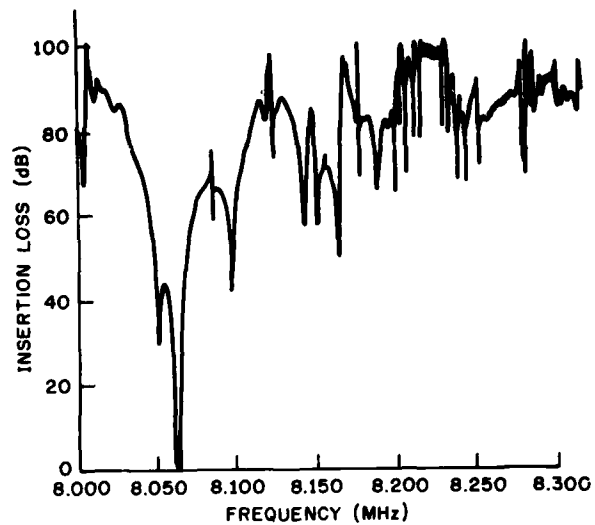


Figure 12
Performance of X-Coupled 8 Resonator Monolithic
Crystal Filter.

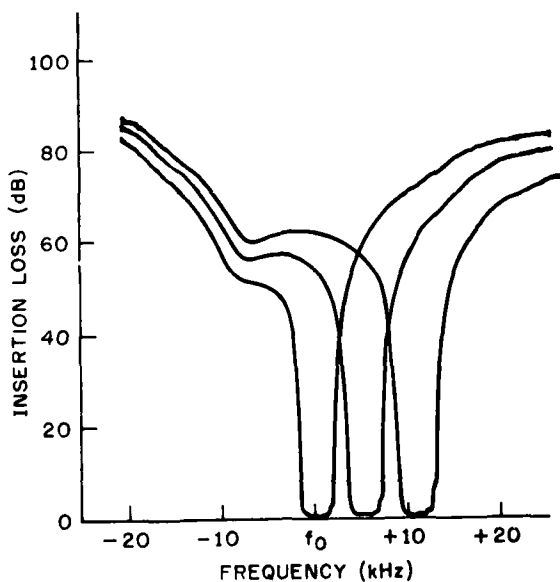


Figure 13
Effect of Changing Filter Frequency by Ion
Milling Electrode Array.

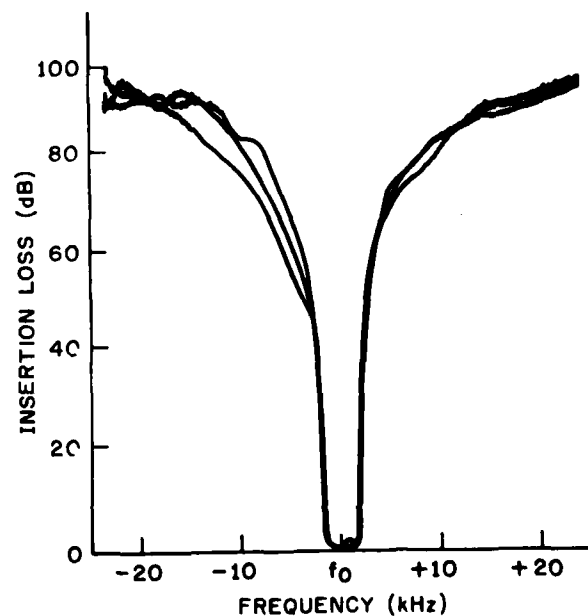


Figure 14
Effect of Changing Plate Length by Lapping
Ends of Plate.

MULTI-MODE STACKED CRYSTAL FILTER

C.M. Stearns, S. Wanuga, S.W. Tehon
General Electric Company
Syracuse, New York

and

A. Kachelmyer
University of Minnesota
Minneapolis, Minnesota

Summary

The stacked crystal filter, a structure where mechanical and piezoelectric coupling occur at the interface of two or more plates, was first introduced in June 1972. The plates are stacked in the thickness direction and are driven in thickness modes of motion. This type of structure may offer miniature low-loss acoustic filters using two or three mode interaction in each crystal layer.

Computer programs have been developed to simulate stacks of two crystal plates with the normal modes arbitrarily located in relation to the plate coordinates. The programs allow the plates to be rotated around the thickness direction with respect to each other.

Computerized filter results will be shown for arbitrarily selected coupling coefficients and mode velocities for the normal modes in the plates. The effects of plate rotation in the case of an AT-cut quartz stack indicates mode coupling for angles as small as 5° with varied response effects for further increases in plate rotation.

In the case of single mode stacks, the effects of the bond parameters on lossy bonds of arbitrary dimensions between plates are treated.

Experimental studies and results of some of the configurations will be presented.

The combined studies indicate that multi-mode stacked crystal filters can be utilized in an effective manner to shape filter response characteristics. They also indicate that characteristics not readily achievable with simple, single-mode units can be obtained.

Introduction

A stacked crystal filter is a bulk wave device composed of individual crystal plates bonded together in a sandwich-type structure. Therefore, the coupling between elements is acoustical (mechanical) rather than the electrical coupling commonly employed in traditional crystal filters. To examine the potential of these stacked filters as miniature low-loss acoustic filters, a series of computer programs were written for the thickness excitation of the thickness mode case. These programs represent single mode plates, plates with two modes normal to the plate thickness direction, and plates with three normal modes arbitrarily located in relation to the plate coordinates.

In order to make this latter program applicable to any generalized configuration, three auxiliary programs are utilized: a program to calculate the plate material constants (stiffness, piezoelectric stress, and dielectric permittivities) for any arbitrary plate orientation in relation to the standard (XYZ) crystal-

lographic axes of the plate material, a second program to determine the orientation of the normal mode axes of the plate in relation to the actual plate coordinates and a third program to convert these results into velocity and coupling coefficient information.

All the results to be described pertain to the simple two-element stack illustrated in Figure 1, however, programs for more elaborate stacks can be developed.

Discussion

The point for the development of these programs was the transmission line normal mode equivalent circuit of Figure 2 as developed by Dr. Ballato [1]. Each transmission line is associated with one of the three allowed modes of vibration as determined from the solution of the secular equation [2] governing this TEM case. If the particular plate under consideration has a mode which is not piezoelectrically excitable, the electromechanical transformers associated with this mode (transmission line) are removed. This normal mode equivalent circuit for each plate can be associated with a normal mode open circuit impedance matrix of the form shown in Figure 3. $Z^{(i)}$, θ_i , and n_i in this figure are respectively the characteristic impedance, propagation constant and electromechanical transformer turns ratio of the allowed mode i . These modes are propagating in the thickness direction of a plate of thickness $2h$. C_0 represents the usual interelectrode capacitance.

If the normal mode axes coincide with the actual plate axes for each element in the stack, this equivalent circuit and matrix contains all the information required to determine the frequency response of the stack, except for the mechanical and electrical boundary conditions. However, in general, the normal mode axes of a plate will not coincide with the actual plate axes. In this case, it is necessary to add a multiple-winding transformer at each end of the normal mode equivalent circuit to represent the orthogonal transformation from the normal mode coordinates to the actual coordinates.

The equivalent circuit representation of a single-element in the stack in terms of the actual plate coordinates is shown in Figure 4. The turns ratios of these transformation transformers are determined from the components of the eigenvectors associated with the eigenvalues of the secular equation. When these same coordinate transformations are applied to the normal mode impedance matrix, the actual coordinate impedance matrix shown in Figure 5 is obtained. In the general case, every term in this actual coordinate matrix is non-zero even though almost half the elements in the normal mode impedance matrix were zero.

The multi-mode filter stacks to be considered here

will consist of only two of these plates as shown in Figure 6. While not shown explicitly in this figure, it is assumed that each plate surface normal to the thickness direction has an electrode on it and that the smallest of these electrodes determines the active area of the plate. There is no requirement that the plates be the same size or thickness or even made of the same material even though the case illustrated uses identical plates. As shown in this figure, another degree of freedom is available in this stack. The actual coordinates of one plate can be rotated about the common thickness axis in relation to the other plate.

The actual mode equivalent circuit of this two-element stack is shown in Figure 7. The multi-winding transformer between the two plates establishes the coordinates of the left plate as the reference framework and refers the coordinates of the right plate to them. The turns ratios are dependent on the relative rotation between the plates. To obtain a solution, the following method is utilized. The traction free mechanical boundary conditions are applied to the free surfaces of both the right and left plates. This reduces the 7×7 impedance matrices to a set of 4×4 matrices involving only the electrical quantities and the equivalent mechanical variables on the junction sides of the plates. Application of the appropriate electrical termination to the right plate further reduces the matrix of this plate to one involving only the junction variables. When the variables of the right plate are referred to the reference coordinates of the left plate, the continuity of stress and velocity conditions can be used to obtain a simultaneous set of expressions for the junction currents in terms of the input electrical current. This leads to an expression for the terminated input impedance of the stack. With an expression for the input impedance, it is a simple task to determine the input current in terms of the applied electrical voltage. The steps are then retraced to determine the appropriate electrical output voltage and current.

This procedure has been applied to several special cases of this generalized equivalent circuit. The first, with only one normal mode perpendicular to the thickness direction in each plate, and no inter-plate rotation, assumes a finite bonding layer between plates. The bond is represented by a lossy transmission line, and has arbitrary values of thickness and cross-sectional area. This case can also be represented by the more familiar Mason's equivalent circuit and solved by straightforward ladder techniques. Figure 8 gives the results computed for two single mode, AT-cut, quartz plates separated by a gold bonding layer with an assumed mechanical Q of 100. Solutions are shown for seven different values of bond area.

Inspection of Figure 8 reveals a direct analogy of the system to a double tuned filter circuit. With full bond area, a single resonance is observed at the wavelength thickness of the stack. With reducing bond area, the single resonance becomes double humped, with ripple and peak separation determined by bond area. Critical coupling is theoretically achievable.

A similar calculation, for fixed bond area and variable bond thickness, shows a similar control of coupling. A thin bond produces the single resonance of loose coupling and a thick bond produces an over-coupled response, accompanied by excessive insertion loss. In all combinations, the optimum critical coupling parameters are dependent on the physical properties of both the bonding material and stack elements.

The second case considered allows two normal modes

in each plate which are again perpendicular to the plate thickness direction. For this case, intimate or welded contact between the mating plate surfaces was assumed, however, relative rotation between the plates was allowed. Figure 9 illustrates the response obtained for two identical plates. The density and dielectric constant of AT-cut quartz material were used as well as a mode 1 coupling coefficient, KK_1 , of 0.088 and a mode 1 velocity, CT_1 , of 3.32×10^3 m/sec. Mode 2 coupling KK_2 , is assumed to be twice that of mode 1, while mode 2 velocity, CT_2 , is 1.1 times that of mode 1. Each plate is resonant and one-half wave thick at 10 MHz for the mode 1 velocity. Assumed electrode diameter is 1 cm, the source is a 1 volt generator with 1600 ohms internal resistance, and the load is a 1600 ohm resistor.

The interaction of these two piezoelectrically coupled modes at 0° of plate rotation results in the notch in the response curve following the mode 1 response. The location of this dip is dependent on the ratio of the mode velocities. A faster mode 2 velocity moves the dip further away from the mode 1 response. At small angles of rotation between plates, the response curve remains essentially unchanged. However, at 45° of relative plate rotation, the mode 1 response has decreased considerably and the notch has become shallower.

Figure 10 shows the effects of a continued rotation of these plates. With 70° rotation between plates, the notch has deepened again, the mode 1 response has disappeared and the mode 2 response has begun to take on a double-humped characteristic. As the relative plate rotation is changed to 80° , the response takes on a skewed characteristic with an approximate mid-band insertion loss of 5 dB, a ± 2 dB passband ripple and a 3 dB bandwidth of approximately 2.5 percent.

When these plates are rotated to 85° , the passband ripple has disappeared as shown in Figure 11. This response has a mid-band insertion loss of 3.5 dB and a 3 dB bandwidth of approximately 1.5 percent.

Figure 12 again shows another stack of two identical plates. The parameters of these plates are the same as those described in conjunction with Figure 9, except for the mode 2 velocity. For this case, the mode 2 velocity is assumed to be slower than the mode 1 velocity and equal to 0.95 times it. This results in the mode 2 response occurring at a lower frequency than that of mode 1. As illustrated for 5° of relative plate rotation, this results in the null response being below the mode 1 response. At a relative plate rotation of 83° , a single peaked response curve is obtained similar to that shown at 85° in the previous case. However, for these conditions, the notch is now above the main response. This immediately suggests the possibility for electrically cascading, simple, two-element stacked filters to tailor overall response. While not illustrated, the relative rotation of the plates can also be used to couple energy into non-piezoelectrically coupled modes in one or both plates. This has the effect of creating notch frequencies in the response curve at the resonant frequencies of these modes. The notch depth and exact character can be controlled by the amount of energy coupled into the mode and is dependent on whether a mechanism to couple any of it back to the electrical output exists or not.

The third case programmed was the 3 mode example illustrated in Figure 7. As mentioned earlier this general case either requires previous knowledge of the plates used in the stack or the use of the three auxiliary programs. Figure 13 illustrates the results

of this program for two AT-cut quartz crystal plates in intimate contact. These plates have one pure piezoelectrically excitable shear mode, a quasi-shear and a quasi-longitudinal mode which are both non-piezoelectrically coupled. With the plate axes exactly lined up (0° relative rotation), the response curve exhibits a nice smooth character. As soon as any relative rotation between the plates is introduced, plate boundary coupling to the other modes takes place and nulls and peaks occur in the response curve.

Figure 14 shows the computed response for a relative rotation of 45° between the plates. Here an interchange of energy from mode 1 to mode 2 occurs. The resonant effects of mode 2 appear and are coupled back to the output through mode 1. The response curve for a relative rotation of 75° between plates is also shown.

The actual response curve for a two-element stack of AT-cut quartz elements with approximately 75° of relative rotation between plates is shown in Figure 15. This curve was taken in a 50 ohm circuit with no attempt at impedance matching with plates individually resonant in the neighborhood of 6 MHz. These plates were bonded together with a synthetic polyester adhesive, Lens Bond from Summers Laboratories, Inc. of Port Washington, Pennsylvania. The advantage of this type of bond, which is capable of being dissolved, is that the same set of stacked crystal elements can be used repeatedly for studying the effects of plate rotation on the frequency response. There is no claim that this response represents a useable filter, but it does serve to lend credibility to the computed response curves. There is a high degree of correlation between the computed response of Figure 14 and the response shown in Figure 15 over comparable frequency ranges.

Conclusion

During this investigation of multi-mode stacked crystal filters the major emphasis was placed on the development of computer programs, based on the transmission line model of a TETM plate, to simulate the stack behavior. This placement of emphasis is the reason why most of the results presented employ identical plates. The programs can handle any combination of plates; however, for debugging and interpretation purposes it is convenient to use identical plates. It is not intended to imply that this configuration is better than or even as good as, any of the multitude of non-identical plate configurations which can be conceived.

Tables I and II indicate some of the wide range of crystalline materials/orientations and bonding materials which can be used in stacked crystal filters. Such a multitude of choices should offer a wide variety of filter characteristics.

It is hoped that enough information about the stacked crystal filter has been presented to demonstrate its potential and to encourage further theoretical and experimental investigation of this configuration. Even the two-element stack considered here deserves more extensive investigation. The next logical step in the investigation after this is the inclusion of a bonding layer in the two-element stack of multi-mode plates. Then an extension of the stack to more than two plates should be carried out.

In conjunction with this extension, an augmentation of the plate equivalent circuit is desirable. This would allow a variety of electrical interconnections between elements in the stack to be considered, and add to the potential usefulness of the

stacked crystal filter.

Acknowledgement

The authors would like to acknowledge the U.S. Army Electronics Command, Fort Monmouth, New Jersey for the support of this program under Contract No. DAAB07-76-C-1337.

References

- [1] A.D. Ballato, "Transmission-Line Analogs for Piezoelectric Layered Structures," PhD Dissertation, Polytechnic Institute of Brooklyn, 1972, also U.S. Army Electronics Command Report ECOM-4413, May 1976.
- [2] H.F. Tiersten, "Thickness Vibrations of Piezoelectric Plates," J. Acoust. Soc. Amer. Vol. 35, No. 1, January 1963, pp. 53-58.

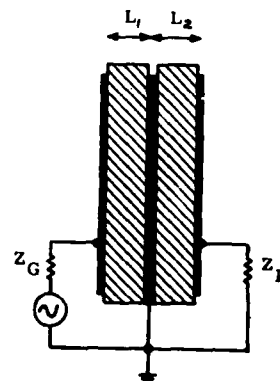


Figure 1 Two-Element Stacked Crystal Filter

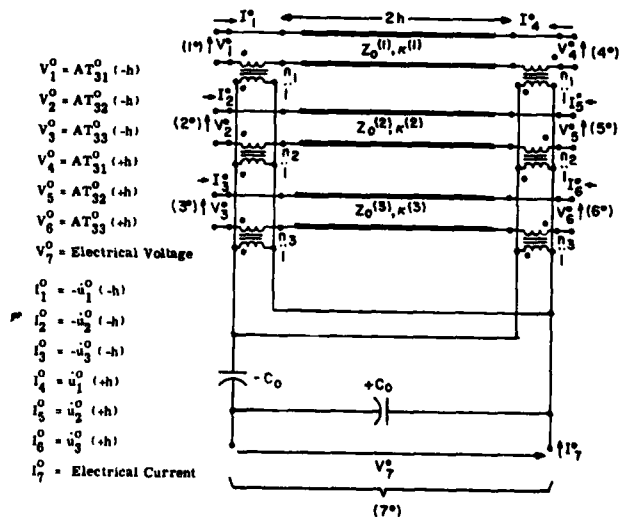


Figure 2 Seven-Port Normal Mode Equivalent Circuit for a TETM Plate of Thickness 2h

$$[Z^0] = \begin{bmatrix} \frac{Z_0^{(1)}}{j \tan \theta_1} & 0 & 0 & \frac{Z_0^{(1)}}{j \sin \theta_1} & 0 & 0 & \frac{n_1}{j \omega C_0} \\ 0 & \frac{Z_0^{(2)}}{j \tan \theta_2} & 0 & 0 & \frac{Z_0^{(2)}}{j \sin \theta_2} & 0 & \frac{n_2}{j \omega C_0} \\ 0 & 0 & \frac{Z_0^{(3)}}{j \tan \theta_3} & 0 & 0 & \frac{Z_0^{(3)}}{j \sin \theta_3} & \frac{n_3}{j \omega C_0} \\ \frac{Z_0^{(1)}}{j \sin \theta_1} & 0 & 0 & \frac{Z_0^{(1)}}{j \tan \theta_1} & 0 & 0 & \frac{n_1}{j \omega C_0} \\ 0 & \frac{Z_0^{(2)}}{j \sin \theta_2} & 0 & 0 & \frac{Z_0^{(2)}}{j \tan \theta_2} & 0 & \frac{n_2}{j \omega C_0} \\ 0 & 0 & \frac{Z_0^{(3)}}{j \sin \theta_3} & 0 & 0 & \frac{Z_0^{(3)}}{j \tan \theta_3} & \frac{n_3}{j \omega C_0} \\ \frac{n_1}{j \omega C_0} & \frac{n_2}{j \omega C_0} & \frac{n_3}{j \omega C_0} & \frac{n_1}{j \omega C_0} & \frac{n_2}{j \omega C_0} & \frac{n_3}{j \omega C_0} & \frac{1}{j \omega C_0} \end{bmatrix}$$

Normal Coordinate Impedance Matrix of a TETM Plate

Figure 3 Normal Coordinate Impedance Matrix of a TETM Plate

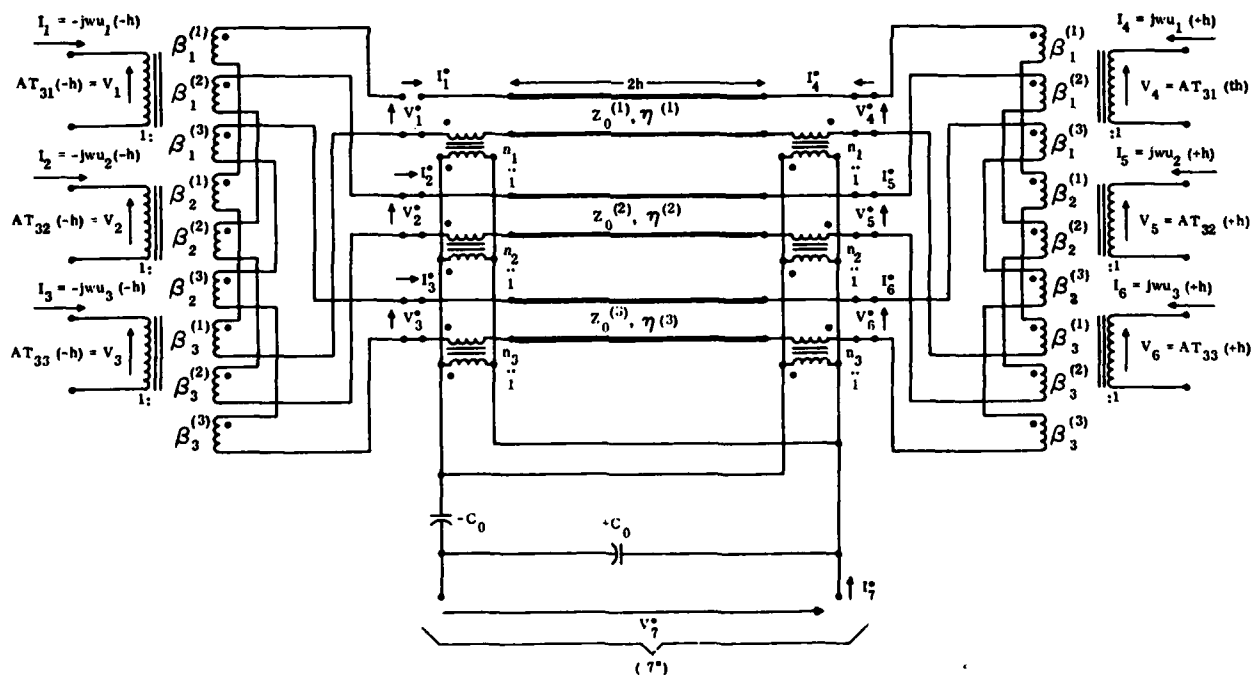


Figure 4 Seven-Port Actual Coordinate Equivalent Circuit for a TETM Plate

$V_1 = AT_{31}(-h)$	$b_{11} Z_{11}^0 b_{11}$ $+b_{12} Z_{22}^0 b_{12}$ $+b_{13} Z_{33}^0 b_{13}$	$b_{11} Z_{11}^0 b_{21}$ $+b_{12} Z_{22}^0 b_{22}$ $+b_{13} Z_{33}^0 b_{23}$	$b_{11} Z_{11}^0 b_{31}$ $+b_{12} Z_{22}^0 b_{32}$ $+b_{13} Z_{33}^0 b_{33}$	$b_{11} Z_{14}^0 b_{11}$ $+b_{12} Z_{25}^0 b_{12}$ $+b_{13} Z_{36}^0 b_{13}$	$b_{11} Z_{14}^0 b_{21}$ $+b_{12} Z_{25}^0 b_{22}$ $+b_{13} Z_{36}^0 b_{23}$	$b_{11} Z_{14}^0 b_{31}$ $+b_{12} Z_{25}^0 b_{32}$ $+b_{13} Z_{36}^0 b_{33}$	$b_{11} Z_{17}^0$ $+b_{12} Z_{27}^0$ $+b_{13} Z_{37}^0$
$V_2 = AT_{32}(-h)$	$b_{21} Z_{11}^0 b_{11}$ $+b_{22} Z_{22}^0 b_{12}$ $+b_{23} Z_{33}^0 b_{13}$	$b_{21} Z_{11}^0 b_{21}$ $+b_{22} Z_{22}^0 b_{22}$ $+b_{23} Z_{33}^0 b_{23}$	$b_{21} Z_{11}^0 b_{31}$ $+b_{22} Z_{22}^0 b_{32}$ $+b_{23} Z_{33}^0 b_{33}$	$b_{21} Z_{14}^0 b_{11}$ $+b_{22} Z_{25}^0 b_{12}$ $+b_{23} Z_{36}^0 b_{13}$	$b_{21} Z_{14}^0 b_{21}$ $+b_{22} Z_{25}^0 b_{22}$ $+b_{23} Z_{36}^0 b_{23}$	$b_{21} Z_{14}^0 b_{31}$ $+b_{22} Z_{25}^0 b_{32}$ $+b_{23} Z_{36}^0 b_{33}$	$b_{21} Z_{17}^0$ $+b_{22} Z_{27}^0$ $+b_{23} Z_{37}^0$
$V_3 = AT_{33}(-h)$	$b_{31} Z_{11}^0 b_{11}$ $+b_{32} Z_{22}^0 b_{12}$ $+b_{33} Z_{33}^0 b_{13}$	$b_{31} Z_{11}^0 b_{21}$ $+b_{32} Z_{22}^0 b_{22}$ $+b_{33} Z_{33}^0 b_{23}$	$b_{31} Z_{11}^0 b_{31}$ $+b_{32} Z_{22}^0 b_{32}$ $+b_{33} Z_{33}^0 b_{33}$	$b_{31} Z_{14}^0 b_{11}$ $+b_{32} Z_{25}^0 b_{12}$ $+b_{33} Z_{36}^0 b_{13}$	$b_{31} Z_{14}^0 b_{21}$ $+b_{32} Z_{25}^0 b_{22}$ $+b_{33} Z_{36}^0 b_{23}$	$b_{31} Z_{14}^0 b_{31}$ $+b_{32} Z_{25}^0 b_{32}$ $+b_{33} Z_{36}^0 b_{33}$	$b_{31} Z_{17}^0$ $+b_{32} Z_{27}^0$ $+b_{33} Z_{37}^0$
$V_4 = AT_{31}(+h)$	$b_{11} Z_{14}^0 b_{11}$ $+b_{12} Z_{25}^0 b_{12}$ $+b_{13} Z_{36}^0 b_{13}$	$b_{11} Z_{14}^0 b_{21}$ $+b_{12} Z_{25}^0 b_{22}$ $+b_{13} Z_{36}^0 b_{23}$	$b_{11} Z_{14}^0 b_{31}$ $+b_{12} Z_{25}^0 b_{32}$ $+b_{13} Z_{36}^0 b_{33}$	$b_{11} Z_{11}^0 b_{11}$ $+b_{12} Z_{22}^0 b_{12}$ $+b_{13} Z_{33}^0 b_{13}$	$b_{11} Z_{11}^0 b_{21}$ $+b_{12} Z_{22}^0 b_{22}$ $+b_{13} Z_{33}^0 b_{23}$	$b_{11} Z_{11}^0 b_{31}$ $+b_{12} Z_{22}^0 b_{32}$ $+b_{13} Z_{33}^0 b_{33}$	$b_{11} Z_{17}^0$ $+b_{12} Z_{27}^0$ $+b_{13} Z_{37}^0$
$V_5 = AT_{32}(+h)$	$b_{21} Z_{14}^0 b_{11}$ $+b_{22} Z_{25}^0 b_{12}$ $+b_{23} Z_{36}^0 b_{13}$	$b_{21} Z_{14}^0 b_{21}$ $+b_{22} Z_{25}^0 b_{22}$ $+b_{23} Z_{36}^0 b_{23}$	$b_{21} Z_{14}^0 b_{31}$ $+b_{22} Z_{25}^0 b_{32}$ $+b_{23} Z_{36}^0 b_{33}$	$b_{21} Z_{11}^0 b_{11}$ $+b_{22} Z_{22}^0 b_{12}$ $+b_{23} Z_{33}^0 b_{13}$	$b_{21} Z_{11}^0 b_{21}$ $+b_{22} Z_{22}^0 b_{22}$ $+b_{23} Z_{33}^0 b_{23}$	$b_{21} Z_{11}^0 b_{31}$ $+b_{22} Z_{22}^0 b_{32}$ $+b_{23} Z_{33}^0 b_{33}$	$b_{21} Z_{17}^0$ $+b_{22} Z_{27}^0$ $+b_{23} Z_{37}^0$
$V_6 = AT_{33}(+h)$	$b_{31} Z_{14}^0 b_{11}$ $+b_{32} Z_{25}^0 b_{12}$ $+b_{33} Z_{36}^0 b_{13}$	$b_{31} Z_{14}^0 b_{21}$ $+b_{32} Z_{25}^0 b_{22}$ $+b_{33} Z_{36}^0 b_{23}$	$b_{31} Z_{14}^0 b_{31}$ $+b_{32} Z_{25}^0 b_{32}$ $+b_{33} Z_{36}^0 b_{33}$	$b_{31} Z_{11}^0 b_{11}$ $+b_{32} Z_{22}^0 b_{12}$ $+b_{33} Z_{33}^0 b_{13}$	$b_{31} Z_{11}^0 b_{21}$ $+b_{32} Z_{22}^0 b_{22}$ $+b_{33} Z_{33}^0 b_{23}$	$b_{31} Z_{11}^0 b_{31}$ $+b_{32} Z_{22}^0 b_{32}$ $+b_{33} Z_{33}^0 b_{33}$	$b_{31} Z_{17}^0$ $+b_{32} Z_{27}^0$ $+b_{33} Z_{37}^0$
V_7	$Z_{17}^0 b_{11}$ $+Z_{27}^0 b_{12}$ $+Z_{37}^0 b_{13}$	$Z_{17}^0 b_{21}$ $+Z_{27}^0 b_{22}$ $+Z_{37}^0 b_{23}$	$Z_{17}^0 b_{31}$ $+Z_{27}^0 b_{32}$ $+Z_{37}^0 b_{33}$	$Z_{17}^0 b_{11}$ $+Z_{27}^0 b_{12}$ $+Z_{37}^0 b_{13}$	$Z_{17}^0 b_{21}$ $+Z_{27}^0 b_{22}$ $+Z_{37}^0 b_{23}$	$Z_{17}^0 b_{31}$ $+Z_{27}^0 b_{32}$ $+Z_{37}^0 b_{33}$	Z_{77}^0

Figure 5 Matrix Relating Actual Coordinate Voltage Variables to Actual Coordinate Current Variables

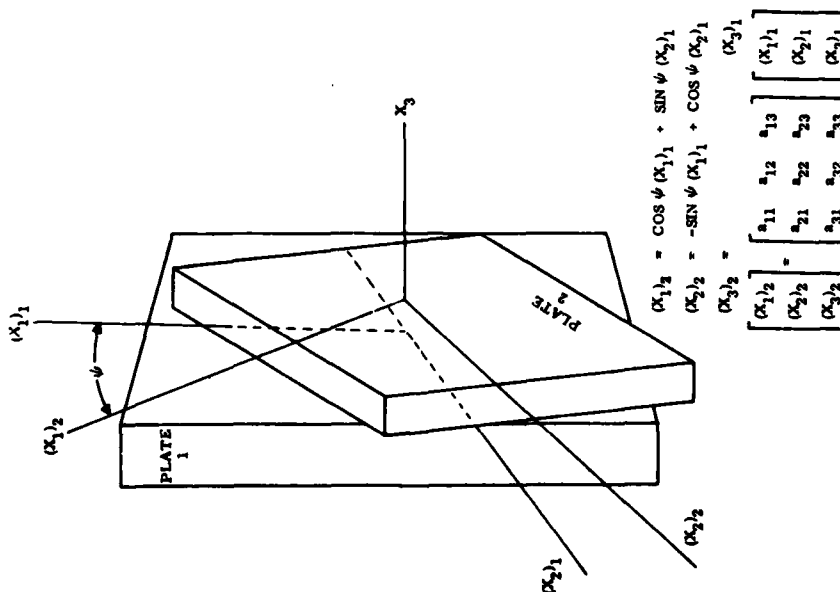


Figure 6 Two-Layer Stack of Crystal Plates Showing Relative Rotation About Common x_3 Axis

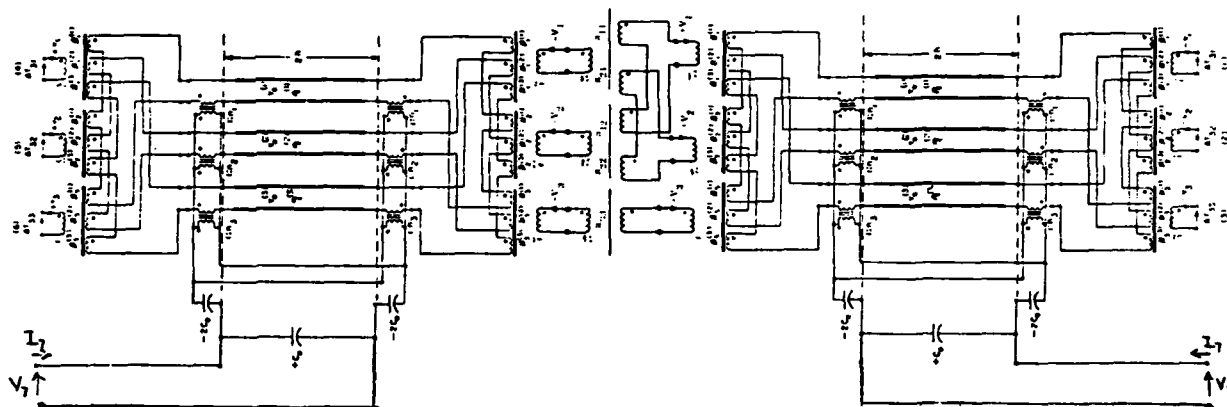


Figure 7 Equivalent Circuit of a Stacked Filter of Two Plates Without Boundary Conditions Applied

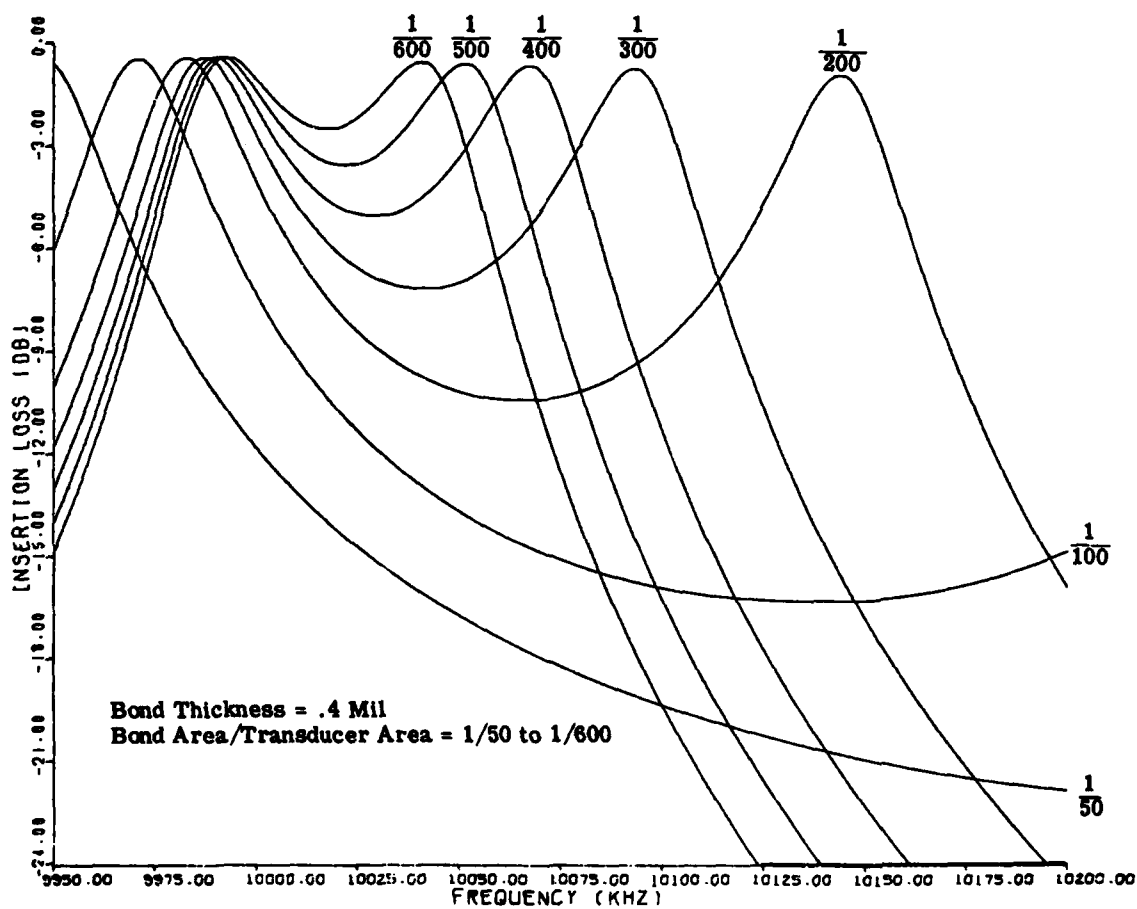


Figure 8 Insertion Loss Versus Frequency for Gold-Bonded Two-Crystal Single-Mode AT-Cut Quartz Filter

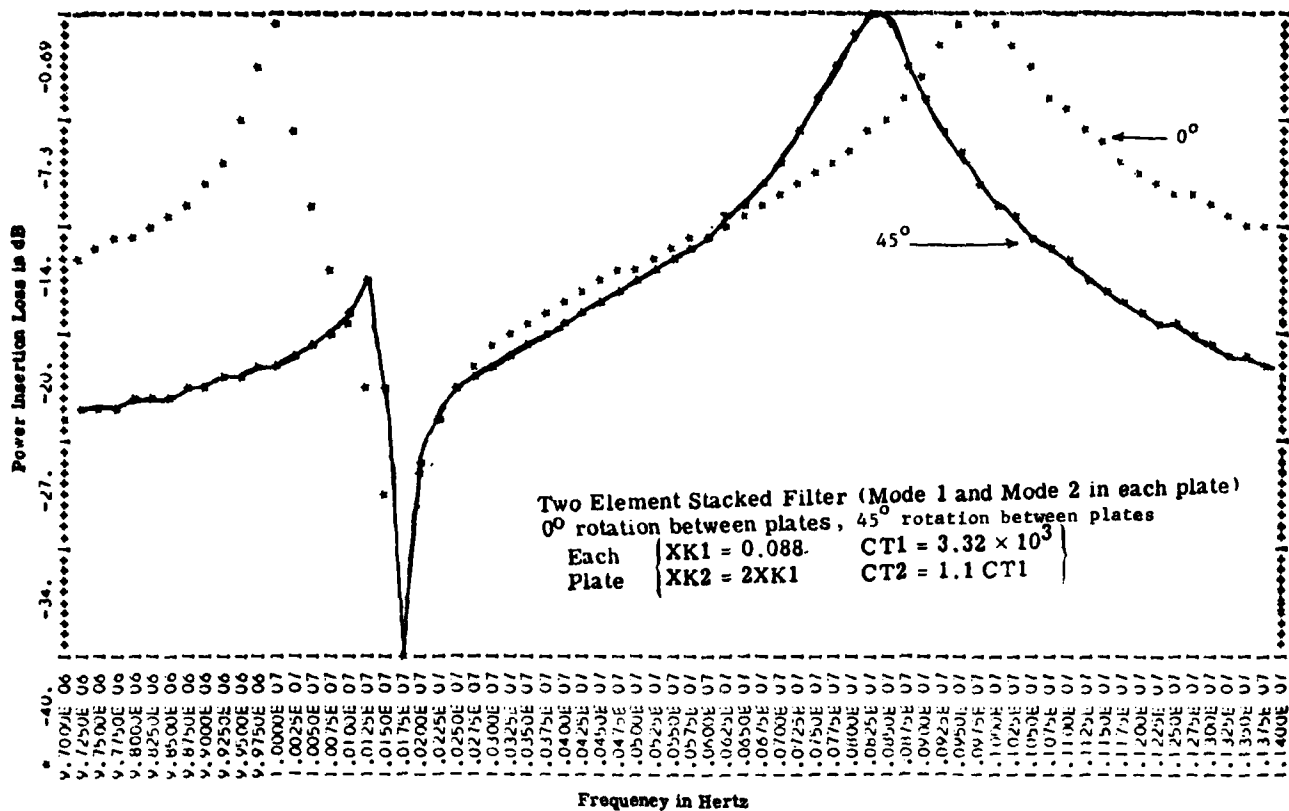


Figure 9 Frequency Response of a Two-Mode Stacked Filter at 0° and 45° of Relative Plate Rotation

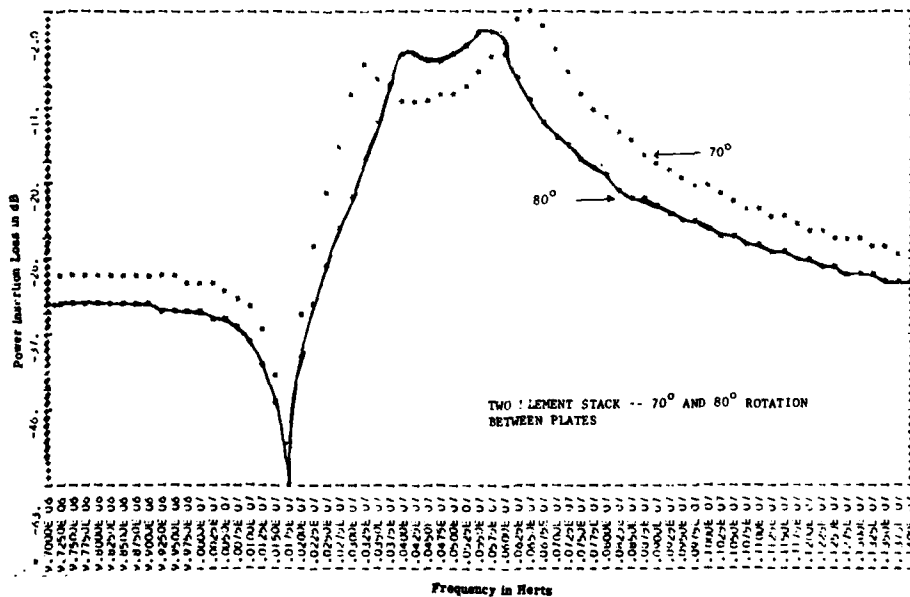


Figure 10 Frequency Response at Relative Plate Angles of 70° and 80° of the Two-Mode Two-Element Stack

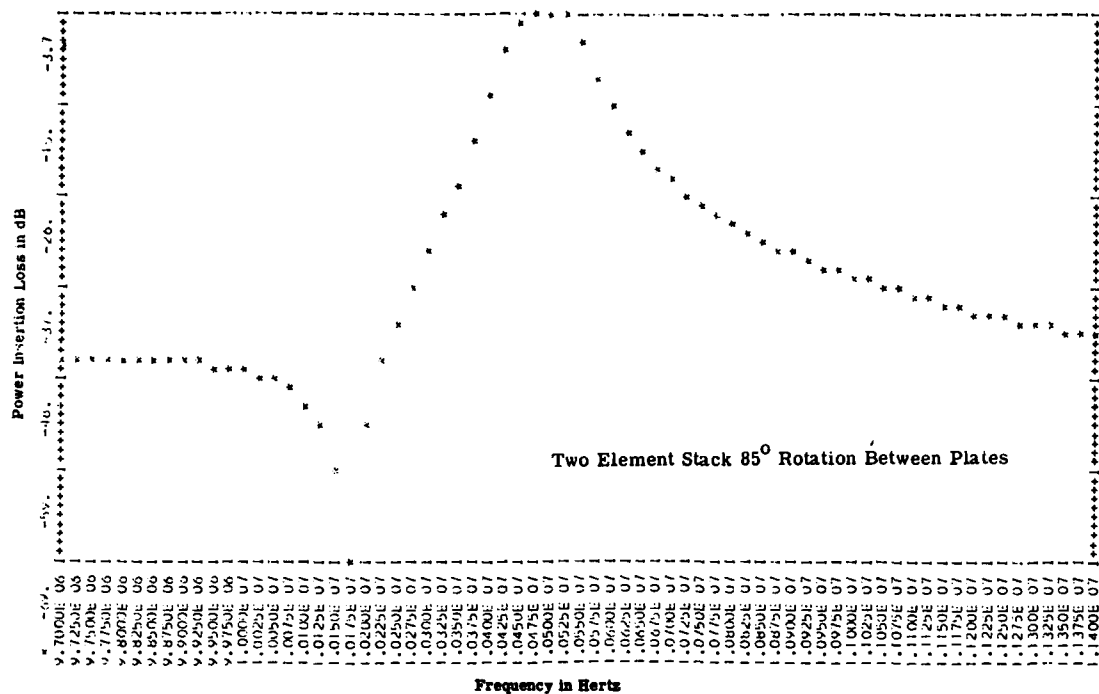


Figure 11 Same Stack of Two-Mode Elements with 85° Rotation Between Plates

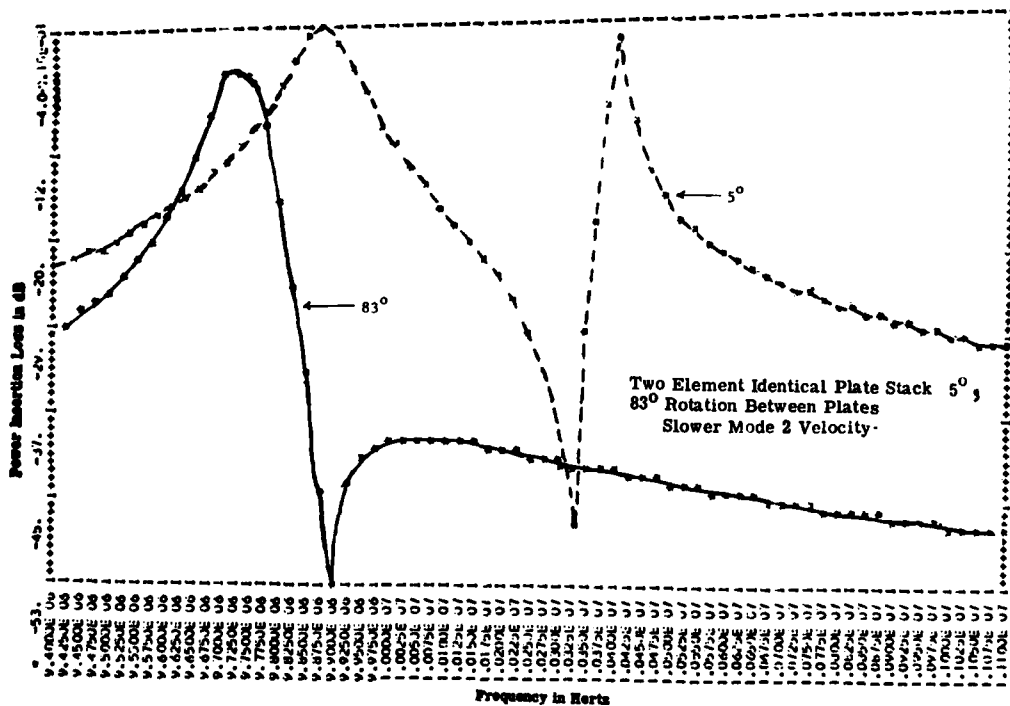


Figure 12 Another Two-Mode Two-Element Stack Frequency Response

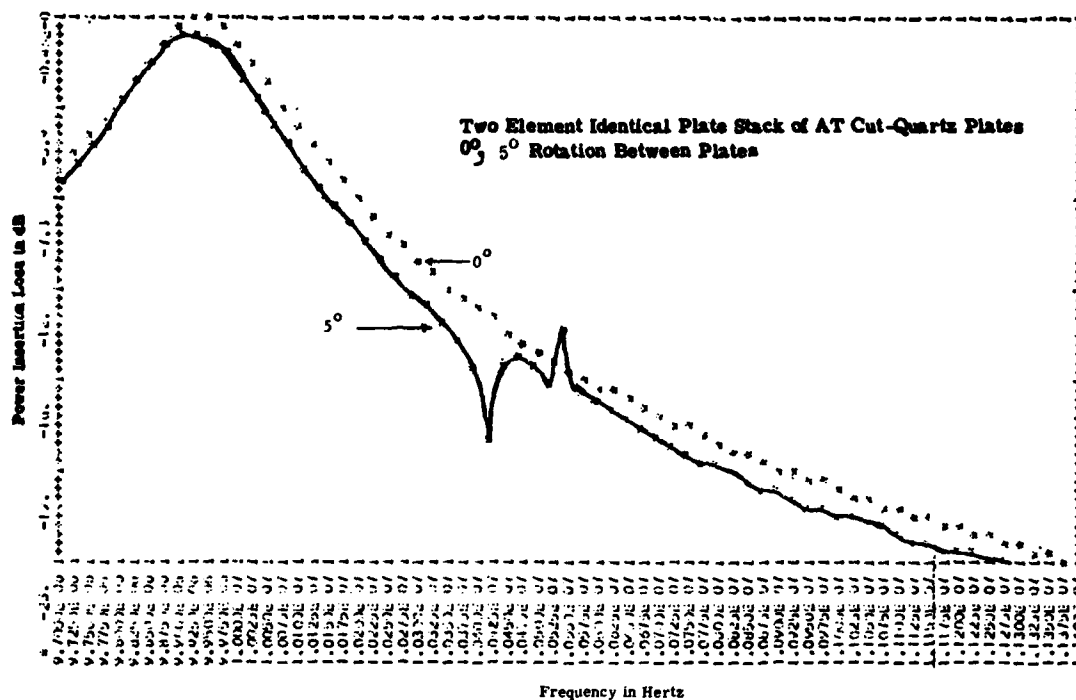


Figure 13 Two-Element Identical Plate Stack of AT-Cut Quartz Plates at 0° and 5° Rotation Between Plates

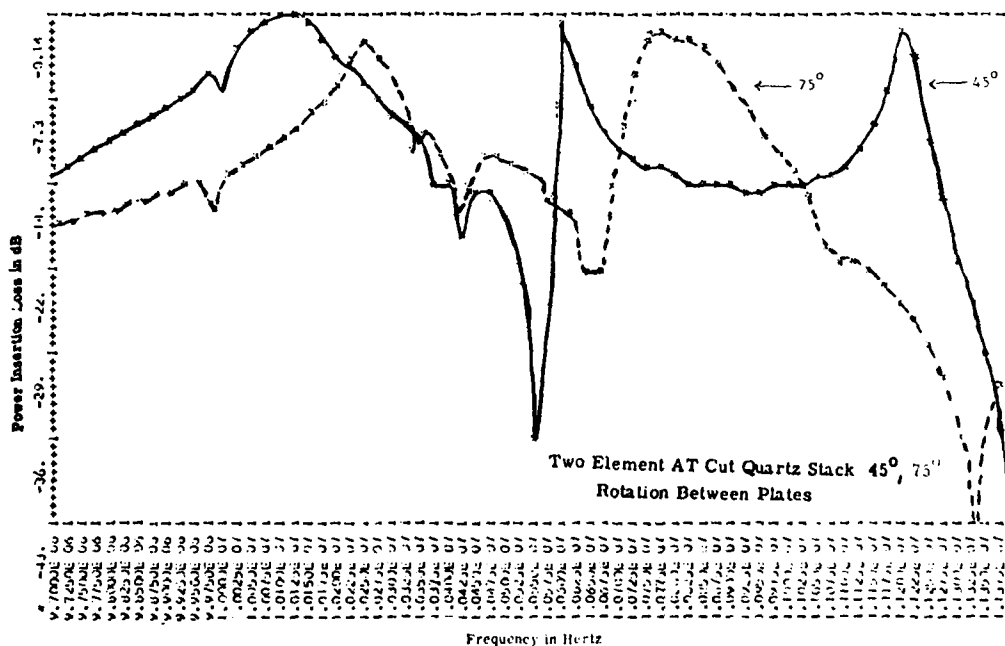


Figure 14 Theoretical Frequency Response of AT-Cut Quartz Stack at 45° and 75°

TABLE I
PROPERTIES OF SOME FILTER MATERIALS

Material	Cut	Mode	k	Temp. Coeff. $10^{-6}/^{\circ}\text{C}$	Vel. $\times 10^3$ m/s	Density g/m^3
Quartz	X	TE	.10	-20	5.70	2.65
Quartz	Y	TS	.14	+90	3.85	2.65
Quartz	AT	TS	.088	0	3.32	2.65
LiNbO ₃	X	TS	.68		4.7	4.7
LiNbO ₃	Z	E	.17			4.7
LiNbO ₃	35°Y	E	.49			4.7
LiNbO ₃	165°Y	S	.62		3.56	4.7
LiTaO ₃	X	TS	.44			5.3
LiTaO ₃	Z	E	.19			5.3
LiTaO ₃	47°Y	E	.29			5.3
LiTaO ₃	165°Y	S	.41			5.3
Ba ₂ NaNb ₅ O ₁₅	X	S	.21		1.82	5.3
Ba ₂ NaNb ₅ O ₁₅	Y	S	.25		1.83	5.3
Ba ₂ NaNb ₅ O ₁₅	Z	E	.57		3.07	5.3

TABLE II
PROPERTIES OF BONDING MATERIALS

Material	Velocity $\times 10^5$ m/sec			$Z \times 10^6$ kg/sec m ²		
Epoxy	2.60 Ext	1.22 Shear		2.86 Ext	1.34 Shear	
Phenyl Benzozate				3.38	0.48	
Indium	2.30	1.44		17.0	10.5	
Gold	3.24	1.20		62.5	23.2	
Silver	3.65	1.61		38.0	16.7	
Copper	5.01	2.11		40.6	18.3	
Aluminum	6.42	3.04		17.3	8.2	
Quartz	5.70	3.32		15.1	8.8	
LiNbO ₃	6.55	4.76		30.8	22.4	
LiTaO ₃	5.55	4.21		29.4	22.3	
BaNaNb ₅ O ₁₅	3.07	1.82		16.3	9.65	

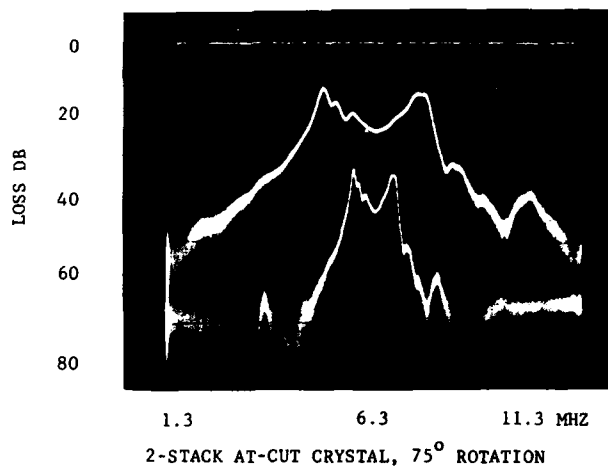


Figure 15

Measured Response of Two AT-Cut Quartz Plates at 75° Rotation Between Plates

128 kHz POLE-TYPE MECHANICAL CHANNEL FILTER

K. YAKUWA, S. OKUDA, K. SHIRAI and Y. KASAI
FUJITSU LIMITED

1015 Kamikodanaka, Nakahara-Ku, Kawasaki, 211, JAPAN

Summary

The design and the characteristics of a newly developed 128 kHz pole-type mechanical channel filter are described.

The filter consists of two 2nd mode flexural transducers, six fundamental mode torsional resonators, longitudinal coupling wires and two two-resonator bridging wires. Since each two-resonator bridging wire is designed to be less than one-half of an acoustic wave-length long, a 1:-1 phase inverter can be realized by making use of these modes of transducers and resonators, without using oblique bridging wire. Then a symmetric pair of attenuation poles above and below the filter pass-band can be realized by each two-resonator bridging wire, respectively. This symple filter structure is well suited for automated production.

Concerning filter synthesis, the Min-Max optimization method is applied to realize the filter circuit having minimized sensitivity to various parameters.

Typical pass-band response fulfills the desired specification of 1/20 of the CCITT recommendation throughout the temperature range of 5°C to 55°C. By the introduction of pole-type mechanical filter, the minimum group delay is reduced to 0.57 ms. The group delay distortion is nearly 1/4 of the CCITT₃ for C 232C. The filter volume is reduced to 10.7 cm³ and is about 1/6 of the conventional LC channel filter now used in FUJITSU.

Introduction

The channel filter for FDM telephone system is the most integral part with the highest production volume in the system. Various filter technologies including LC filters, mechanical filters, MCFs etc. have been studied for improving cost/performance ratio for use in next generation of telephone system.

In the field of mechanical filters, two types of filters, polynomial and pole-type, have been studied.

As for polynomial mechanical filters, a compact channel filter at 48 kHz has been developed by Siemens, Germany. The stop-band selectivity and the pass-band amplitude characteristics of this filter are satisfactory for voice communication, but the group delay distortion may be excessive when data are to be transmitted.

On the other hand, pole-type mechanical filter³⁻⁸ allows the reduction in number of resonators for the same stop-band selectivity, thereby giving a flatter group delay in pass-band and a smaller size.

The authors realized a pole-type mechanical channel filter at 128 kHz in a simple structure for a premodulation scheme use where only one filter type is required.

Filter Structure

The pictorial structure of this filter is shown in Fig.1. Transducers and resonators are coupled by longitudinally vibrating wires. Two two-resonator bridging wires are used to realize two pairs of attenuation poles at finite frequency in the upper and the lower stop-band. In this figure, arrow heads and a dotted line indicate the direction of the vibration displacements. Displacement direction at A and B are inverted by making use of flexural transducer and torsional resonator. Thus a phase inverter can be realized by bridging these two points by a bridging wire which is less than one-half of an acoustic wave-length long, without using oblique bridging wire.

As the result, a symmetric pair of attenuation pole in the upper and the lower stop-band can be realized by each two-resonator bridging wire. This simplified structure is well suited for automated production. An electric circuit is connected to each of input and output of the mechanical filter, which acts as an impedance matching circuit and an electrical coupling circuit of the signal filter to the channel filter. The equivalent circuit of the filter is shown in Fig.2.

Circuit Design

The mechanical channel filter was designed to minimize the number of resonators for miniaturization and economization, by introducing the attenuation pole and utilizing the selectivity of the electric matching circuits as well as the transducers. From the results, the circuit degree was determined to 20th. To realize a filter circuit having minimized sensitivity to various parameters, sensitivity investigations were performed.

From the results, design condition of 128 kHz pole-type channel filter was determined as follows:

- . reflecting propagation attenuation
in the pass-band: 33.4 dB
(i.e. a ripple = 0.004dB)
- . circuit degree : 20th
- . pole frequency : 127.645 kHz
127.905 kHz
131.790 kHz
132.004 kHz

Concerning circuit synthesis, the Min-Max optimization method was applied.

Resonator and Transducer

Resonator

The fundamental mode torsional resonators are made of a constant Young's modulus alloys, and are approximately 11 mm in length and 3 mm in diameter. After suitable heat treatment, the average Q-value of these resonators becomes about 3×10^4 and the frequency

temperature coefficient is less than $2 \times 10^{-6}/^{\circ}\text{C}$. The resonators are tuned automatically by a laser machine within $\pm 4\text{Hz}$.

Transducer

The highly stable and high-Q transducer was developed for this filter. The transducer is vibrating in 2nd mode flexure, and is made of a constant Young's modulus alloy and a piezo-electric ceramic having a coupling factor k_{31} of 27.5 %. The ceramic platelet is soldered to a flattened side of a resonator as shown in Fig. 3.

The poling directions of the ceramic platelet are inversed as shown in this figure, so that the 2nd flexural mode could be driven. The supporting wire is welded to the center of the transducer. Characteristics of the transducer are as follows:

- . frequency temperature coefficient: $< 5 \times 10^{-6}/^{\circ}\text{C}$
- . mechanical Q: approx. 3500
- . capacitance ratio: 400

Coupling between resonators

The fundamental mode torsional resonators are coupled by longitudinally vibrating coupling wire by means of welding. Coupling state of two resonators is shown in Fig. 4. A coupler is spot-welded at the position X on the two resonators. Where equivalent coupler length is l_c , and D_r , D_c are the diameter of the resonators and the coupler, respectively. Resonance angular frequencies at both resonators are equal to ω_r before coupling.

When only the longitudinal coupling is considered, the equivalent circuit is shown in Fig. 5. In the coupling state, the two observed resonance angular frequencies ω_1 and ω_2 (corresponding to the zeros of the input impedance), coupling coefficient k and center angular frequency ω_0 are given as follows:

$$\omega_1 = \sqrt{S_r / (M_r + M_c)} \quad (1)$$

$$\omega_2 = \sqrt{(S_r + S_c) / (M_r + M_c)} \quad (2)$$

$$\omega_0 = (\omega_1 + \omega_2) / 2 \quad (3)$$

$$k = (\omega_2 - \omega_1) / \omega_r \approx (\omega_2 - \omega_1) / \omega_1 \quad (4)$$

where

$$M_r = \rho_r \cdot A_r \cdot l_r / (2 \zeta_x^2) \quad (5)$$

$$S_r = \omega_r^2 M_r \quad (6)$$

$$M_c = (z_0 / \omega_0) \cdot \tan(\omega_0 l_c / 2 v_c) \quad (7)$$

$$S_c = \omega_0 z_0 \cdot \operatorname{cosec}(\omega_0 l_c / v_c) \quad (8)$$

$$\zeta_x = \sqrt{2} \cdot \cos(X / l_r) \quad (9)$$

$$z_0 = A_c \sqrt{\rho_c E_c} \quad (10)$$

where A_r , A_c , and ρ_r , ρ_c indicate the sectional areas and densities of the resonator and the coupler, respectively, and the v_c , E_c indicate the longitudinal velocity and Young's modulus of the coupler.

Relation between coupling coefficient and coupler length in case of $X=1.03\text{ mm}$ and $D_c=0.218\text{ mm}$ is shown in Fig. 6. "o" represents the measured value and the

solid line represents the value calculated from equation (4). In the measured value, abnormal state of coupling coefficient are found around points of $l_c=4.5\text{ mm}$ and $l_c=6.3\text{ mm}$. These are caused by resonances of the flexural modes of the coupler.¹⁰ The dotted line indicate the calculated value in consideration of the effect of flexural modes, and in this case the calculated value well agrees with the measured one.

Thus it is needed to avoid flexural resonance of coupler when determining dimensions of coupler, and coupler length. By this way, error of coupling coefficient can be reduced to about 1 %.

Characteristics

Fig. 7 shows the external appearance of the filter. The filter volume is 10.7 cm^3 including matching circuits, and is about 1/6 of the conventional LC channel filter now used in FUJITSU.

Typical attenuation characteristics of the filter is shown in Fig. 8. The pass-band response satisfies 1/20 CCITT recommendation throughout the temperature range of 5°C to 55°C . The characteristics is achieved only by the tuning of the end transformers, without adjustment of assembled mechanical part of the filter. The insertion loss at 800 Hz off carrier (i.e. 128.8 kHz) is about 1.4 dB.

Group delay characteristic is shown in Fig. 9. The minimum group delay is reduced to 0.57 ms. The group delay distortion is nearly 1/4 of the CCITT Rec. G.232C.

Fig. 10 shows the spurious responses in the region of 10 kHz to 1 MHz. The spurious response more than 60 dB is obtained by using end electric circuits and making vibration mode of transducers different from that of resonators.

Fig. 11 shows the aging characteristics of the filter under the condition of 55°C . As shown in this figure, characteristics is excessively stable.

Conclusion

A pole-type mechanical filter at 128 kHz has been successfully developed in a very simple structure, which is well suited for automated production.

The number of resonators are minimized to 8 including two transducers, for miniaturization and economization. By applying the minimized sensitivity circuit design, pass-band response fulfills the desired specification of 1/20 CCITT recommendation throughout the temperature range of 5°C to 55°C .

The minimum group delay is also reduced to 0.57 ms., and the group delay distortion is nearly 1/4 of the CCITT Rec. G.232C.

Acknowledgment

The authors would like to express their appreciation to Dr. K. Sawamoto of Yokosuka Electric Communication Laboratory, NTT. for his fruitful discussion, to Dr. M. Kawashima, General Manager of Transmission Division, FUJITSU LIMITED, for his consistent guidance and encouragement.

References

1. H. Albsmeier, "Ein vergleich der realisierungsmöglichkeiten elektromechanischer Kanalfilter im Frequenzbereich 12 kHz bis 10 MHz", Frequenz 25 (1971) 3 pp.74
2. K. Yakuwa, S. Okuda and M. Yanagi, "Development of new channel bandpass filters", Proc. ISCAS/74 pp.100 (1974)

3. H. Schüssler, "Filter for channelbank filtering with mechanical resonators", Proc. ISCAS/74 pp.127(1974)
4. R. A. Johnson and W. a. Wingot, "FDM equipment using mechanical filters", Proc. ISCAS/74 pp.127(1974)
5. K. Yakuwa, S. Okuda, Y. Kasai and K. Asano, " Pole-type mechanical filter", The Committee of Electro-Mechanical Functional Devices, IEE of Japan, 36-88 June 21(1974) (in Japanese)
6. T.Yano, T. Futami and S. Kanazawa, "New torsional mode electromechanical channel filter", 1974 IEEE European Conf. on Circuit Theory and Design. London, July(1974)
7. K. Sawamoto, E. Sasaki, S. Kondo, T. Ashida and K. Shinozaki, "A torsional mode mechanical filter", Trans. IECE of Japan, 57-A, pp.575(1974)
8. K. Sawamoto, S.Kondo, N. Watanabe, K. Tsukamoto, M. Kiyomoto and O. Ibaraki, " A Torsional Mode Pole-Type Mechanical Channel Filter", IEEE Trans. on SU -23, NO.3, pp.148, May(1976)
9. M. Börner, "Mechanischer filter mit Dampfungspolen" A.E.U., 17, pp.103(1963)
10. S. Sugawara, T. Ogasawara and M. Konno, "Spurious response of torsional mode mechanical filter", Paper of Technical Group on CTS76-69 IECE of Japan Oct. 20(1976)

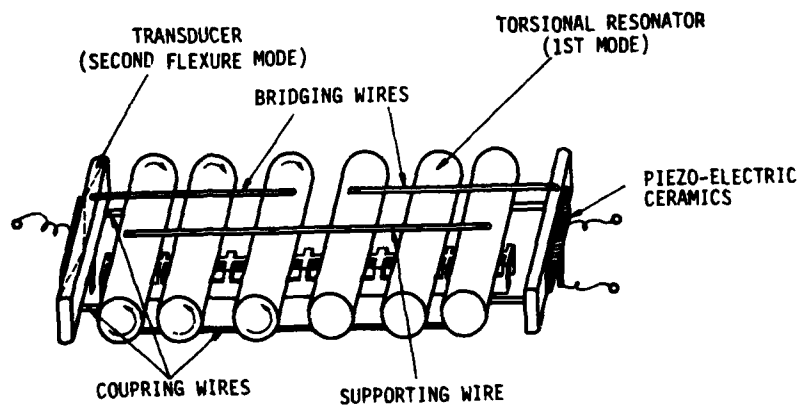


Fig.1 Structure of Mechanical Channel Filter

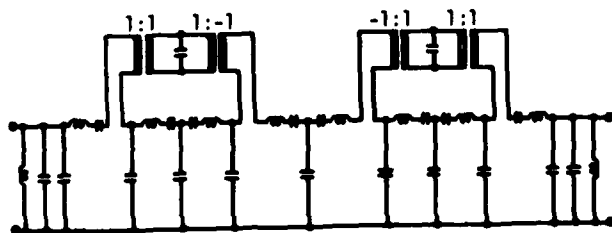


Fig.2 Equivalent Circuit

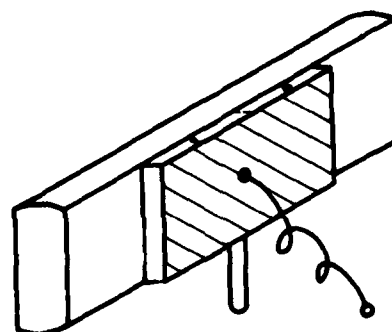


Fig.3 2nd Mode Flexural Transducer

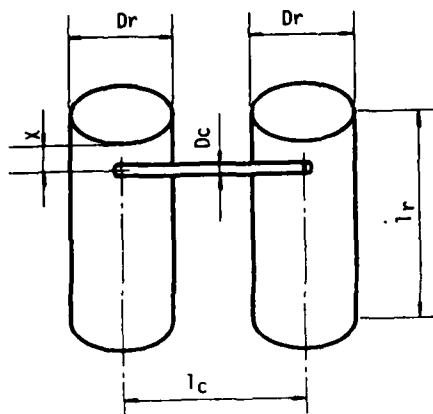


Fig.4 Coupling State between Two Resonators

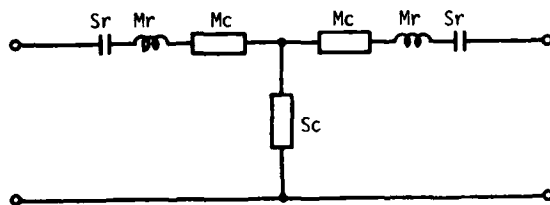


Fig.5 Equivalent Circuit of Coupled Resonators

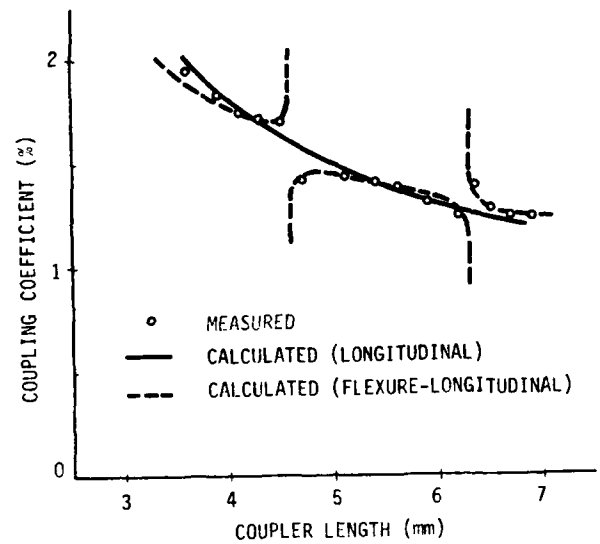


Fig.6 Coupling Coefficient vs. Coupler Length

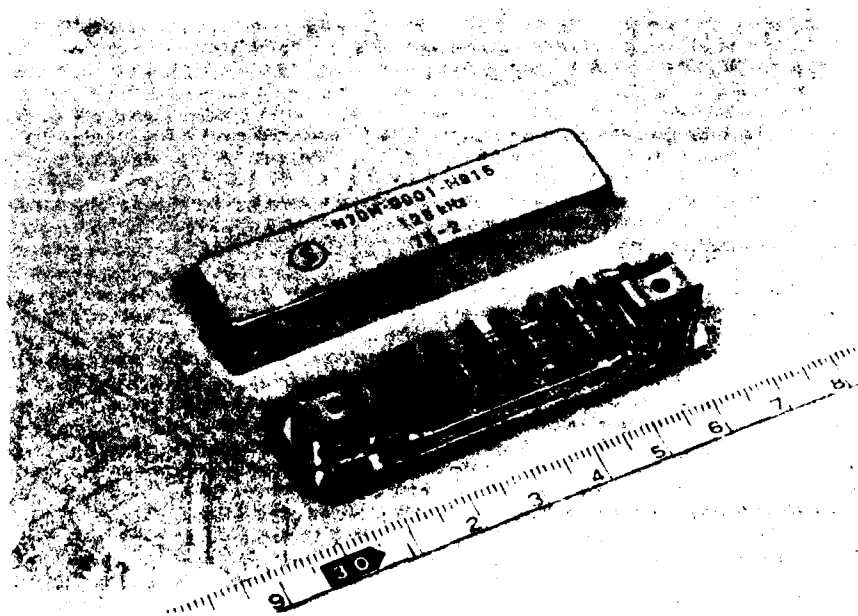


Fig.7 128 kHz Pole-Type Mechanical Filter

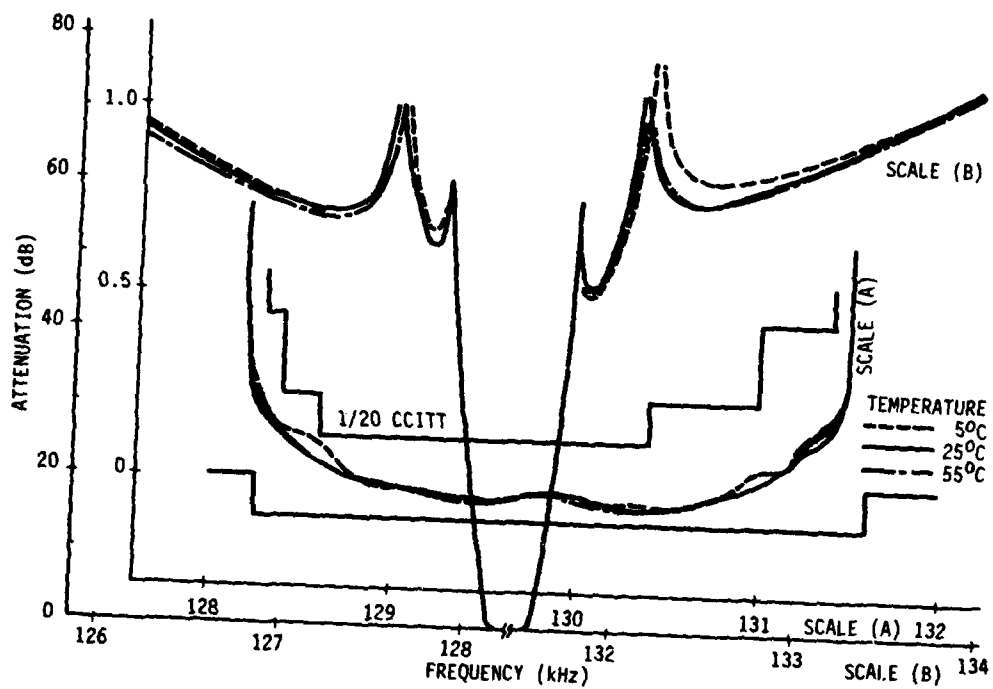


Fig.8 Temperature Characteristics of 128 kHz Pole-Type Mechanical Filter

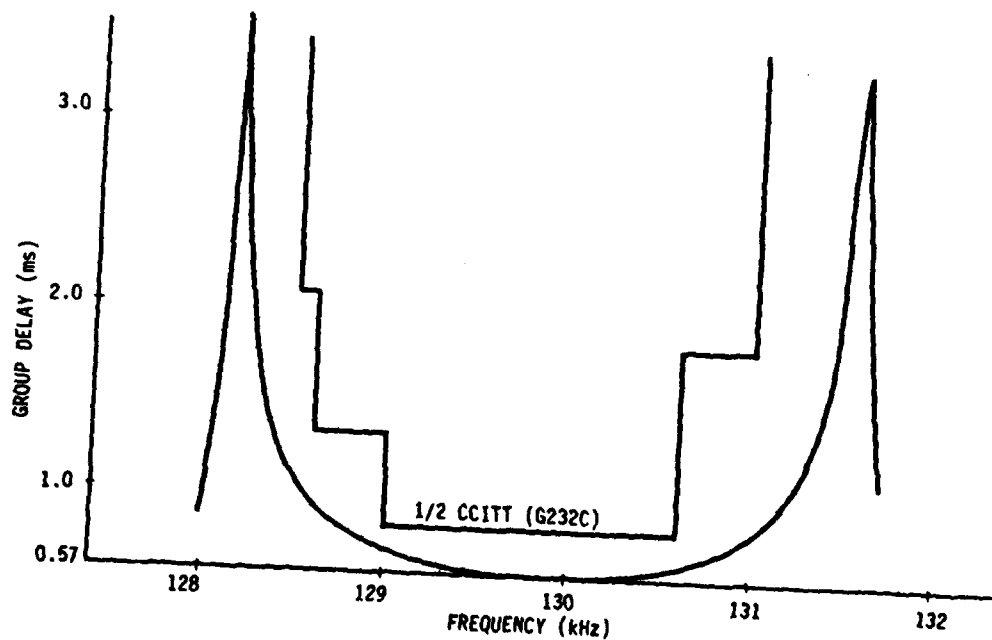


Fig.9 Group Delay Characteristic

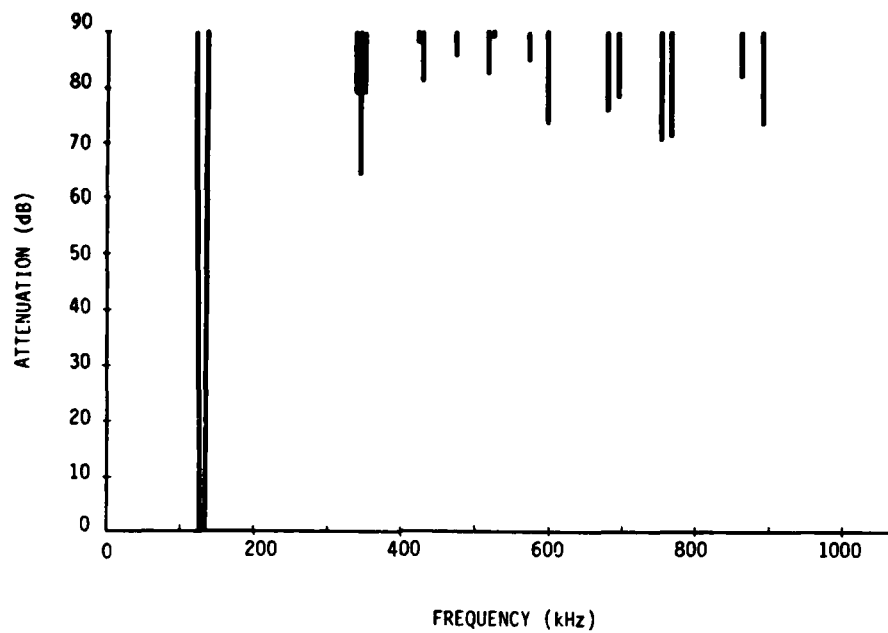


Fig.10 Spurious Response

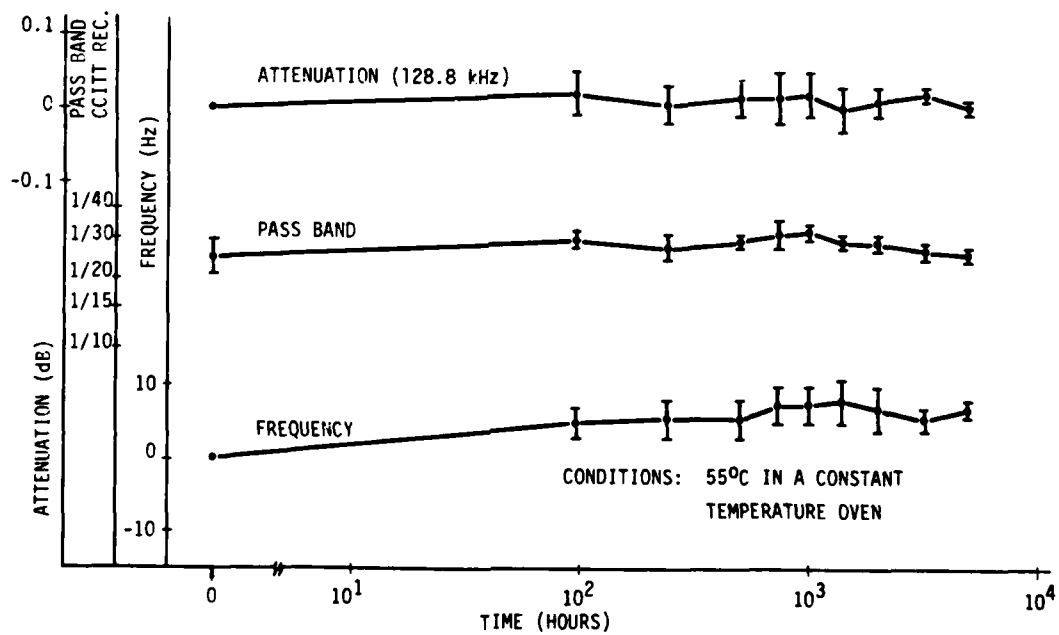


Fig.11 Aging Characteristics

LITHIUM TANTALATE CHANNEL FILTERS FOR MULTIPLEX TELEPHONY

Tomas Arranz

Compagnie d'Electronique et de Piézoélectricité
78500 - Sartrouville - France

Summary

There exist various possibilities provided by modern filtering techniques for the realization of channel filters for multiplex telephony ; a new procedure, based on the use of Lithium Tantalate resonators, has been studied in detail. It turns out that by using new piezoelectric materials, such as Lithium Tantalate, it is possible to extend the range of applications of crystal filters into a domain that was dominated previously by mechanical and by classical LC filters.

The present paper describes in detail the successive stages involved in the realization of a Lithium Tantalate channel filter for 128 KHz. Its characteristics correspond to the CCITT standards.

The scheme of the paper is as follows :

- a succinct description of the various methods currently in use : L-C filters, mechanical filters, quartz filters
- design and optimization of a Lithium Tantalate filter
- design and realization of Lithium Tantalate resonators
- realization of a filter
- critical analysis of the results and conclusions.

Introduction

During the past few years, work on new multiplex telephony systems has taken on added importance, since the present system (premodulation at 24 KHz with LC filtering) may be improved and therefore should be revised. Due to different premodulation frequencies, the new systems require a variety of filters ; improved LC filters for 24 KHz, ceramic filters, mechanical filters at 128 KHz or 256 KHz, classical quartz filters at 2.5 MHz or monolithic crystal filters at 8.192 MHz (U.S.A.).

The technology to be adopted for the realization of channel filters is strictly dependent upon the physical realizability of the components. The limitations of the various technologies are reasonably represented in figure 1, ^{1, 2, 3} on which we have superposed the operational domain of Lithium Tantalate (LiTaO₃).

It should be emphasized that the use of Lithium Tantalate allows one to cover a new operational domain which is complementary to that of quartz.

An examination of figure 1 shows that, for frequencies of some kilohertz and for relative bandwidths between 0.4% and 10%, the mechanical filter provides an interesting solution which explains the intensive work being carried out in this field. Furthermore, it is immediately apparent that LiTaO₃ piezoelectric resonators cover essentially the same domain.

In table I, we have assembled the main design parameters of the materials or means that may be used for the realization of channel filters for telephony ; here,

$$q = Q \times \frac{3 \text{ dB bandwidth}}{F_0}$$

Being given that the normalized quality factor, q , is highest for LiTaO₃ and that its frequency-temperature characteristic is acceptable, it seemed reasonable to realize a multiplex telephony channel filter at 128 KHz using this material.

Characteristics of the filter

The tolerance contours that are considered ($\frac{1}{10}$ CCITT) are shown in figures 2 and 3 for the passband and the stopband, respectively. The allowable group delay variations are given in figure 4.

The tolerance contours imply extremely high selectivity for the filter : 40 dB suppression of the carrier and ± 0.22 dB ripple in the passband for a temperature range of $+10^\circ$ to $+60^\circ$ C. Furthermore, the accepted insertion loss must be less than 1.5 dB and the input and output impedances must be identical and equal to 600 Ω . These stringent conditions necessarily lead to a transfer function having a large number of poles.

In order to ensure a simple and easily reproducible structure for the filters using LiTaO₃ resonators and keeping in mind the high attenuation required in the stopband (70 dB), it was decided to use a Cauer transfer function having finite attenuation poles realized as a ladder structure (figure 5).

Filter design procedure

The synthesis procedure used is described in detail in the classical paper of Saal and Ulbrich ⁵ and leads to the low-pass structures shown in figure 6.

Actually, whenever possible, the element values of the normalized low-pass filter are read from the tables of the appendix of reference 5, after having determined the degree of the filter and the selectivity factor (or modular angle) from the tolerance contours. The resulting circuit (or its corresponding transfer function) may then be analysed in order to verify that attenuation and group delay responses lie within the required limits ⁶. For certain critical cases, the entire synthesis procedure may be carried out automatically by means of a specific computer programme ⁴.

For $n = 10$, the normalized low-pass filter is given in figure 7. This circuit is progressively modified by means of standard network transformations ^{5,7,8} in order to approach the desired circuit of figure 5.

The transformations are as follows (figures 8, 9, 10) :

- denormalization relative to bandwidth.
- partition of the network into elementary cells.
- low-pass to band-pass transformation in which negative capacitances appear.
- elimination of the central negative capacitances by means of ideal Transformers.
- elimination of the negative capacitances in the input and output cells (Norton's theorem).
- absorption of the ideal transformers.
- transformation of the central cells into crystal resonator equivalent circuits.
- final denormalization and introduction of the piezoelectric resonators.

The Lithium Tantalate Resonators 9, 10.

The equivalent circuit of a LiTaO_3 resonator is identical to that of quartz (figure 11) ; however the element values differ considerably. Bandwidth considerations lead to a large electro-mechanical coupling coefficient ; consequently, a $Y + 40^\circ$ cut which is suitable for a longitudinal mode was adopted (Figure 12).

The resulting capacitance ratio, $\frac{C_0}{C}$, is approximatively 20 and is related to the coupling coefficient as follows :

$$\frac{C}{C_0} = \frac{8}{\pi^2} \cdot \frac{k^2}{1 - k^2}$$

The resonant frequency of the crystal in the longitudinal mode is given by the relation :

$$F = \frac{1}{2 L_g \sqrt{\rho S_{22}''}}$$

where S_{22}'' is the compliance, ρ the density and L_g the length of the bar. Furthermore, the equivalent inductance is proportional to the thickness of the bar :

$$L = K e \cdot \frac{L_g}{g}$$

The above relations allow one to estimate the pertinent characteristics relative to the resonator, e.g., the quality factor is approximatively 50.000.

Frequency variation as a function of temperature is of prime importance. In the case of LiTaO_3 , the frequency-temperature characteristics for series and parallel resonances are quite different (figures 13a, 13 b).

The frequency variation for series resonance is small ($\approx 20 \text{ Hz}$) and it should be possible to improve this performance by varying the cut angles θ or ψ of figure 14. The parallel resonance problem is entirely different (180 Hz) : compensation is therefore required.

Realization and Optimization of the Filter

The above transformations have led to the circuit shown in figure 10 c which consists of 8 crystal resonators and a series LC circuit. The resulting equivalent inductances of the crystal resonators vary widely : an industrial realization of these resonators, per se, is precluded. Consequently, four standard inductances were chosen : 1.1 H, 0.8 H, 0.4 H and 0.2 H.

Under these conditions, the results are given in figure 15 and figure 16, the tolerance contours being respected. However, the series LC circuit is undesirable because of its relatively low quality factor and its poor temperature behaviour. It is replaced by means of the transformation indicated in figure 17. The definitive network, consisting of 9 resonators, is given in figure 18. The frequency response characteristics (figures 19 and 20) compare favourably with those obtained with the original filter (figures 15 and 16).

Temperature Compensation

Variations in temperature will, of course, change the frequency response curve of the filter (figure 21). Among the multiple factors that obtain, we shall consider those that seem most pertinent :

Transformers

Temperature variation of inductances may be compensated by means of tuning capacitances having a negative temperature coefficient, e. g., a combination of

$$L = L_0 (1 + a t)$$

$$\text{and } C = C_0 (1 - a t)$$

will yield a satisfactory frequency variation as a function of temperature.

Resonators

The resonator problem is twofold :

- the deviation of the series resonant frequency must be compensated by varying the cut-angle slightly, thereby permitting a translation of the extremum of the frequency- temperature characteristic, e.g., for an inversion point at 35°C , the frequency variation will be of the order of 6 Hz.

- compensation of the deviation of the parallel resonance frequency is necessary and is readily carried out whenever the parallel capacitance of the crystal resonators is relatively high.

The results obtained after temperature compensation are given in figures 22 and 23.

Realization of an operational model

Two solutions were adopted (photograph 2).

External matching

A printed circuit is used for the realization of these filters ; the overall volume is thereby reduced to a minimum. However, the external matching circuits must be realized with care.

Internal matching

In this case, the volume of the filter is increased considerably ; however a suitable volume may be obtained by changing the shape of the transformer.

Evaluation of the results

The performances achieved with the filter described above are not definitive and may be improved in subsequent realizations. In particular, the 128 KHz range does not bring out the full potential of Lithium Tantalate since its preferential operational frequency is approximately 250 KHz.

In fact, the realization of a filter at 250 KHz presents undeniable advantages. The large external shunt capacitance permits a relatively easy temperature compensation of the parallel resonance frequency, the Q-factor is adequate for this frequency and finally the overall dimensions of the filter are reduced by half.

In fine, one may arrive at the following conclusions concerning the use of Lithium Tantalate for channel filters :

- LiTaO_3 filters provide an excellent alternative to the use of mechanical filters
- the overall dimensions of the filters compare favourably with those of comparable mechanical filters
- the realization of LiTaO_3 resonators will benefit from the experience and technology acquired with quartz resonators.
- At present, LiTaO_3 is expensive ; its cost will fall if LiTaO_3 filters are shown to be compatible and competitive with the filters used at present in this domain.

Acknowledgements

The author would like to thank Mr. Bidart and Mr. Kerboull for their help and encouragement ; particular thanks are due to M. Savy for the theoretical and experimental work leading to the realization of the LiTaO_3 resonators.

References

1. M. Battini, F. Caviglia, Filtri elettromeccanici di canali per multiplex telefonici a divisione di frequenza, *Elettronica Tel.* 22,05, Septembre 1973, p.p. 183 - 193.
2. R. Bosc, F. Collombat, P. Loyez, Applications de nouvelles technologies de filtrage à un système à 12 voies, *Cables et Transmissions*, 1, Janvier 1973, p.p. 103 - 125.
3. G. Duval, J. Renaud-Goud, Considérations sur l'étude d'une nouvelle génération d'équipements terminaux analogiques à 12 voies, *Cables et Transmissions*, 1, Janvier 1973, p.p. 82 - 97.

4. C. Gimenes, Synthèse des filtres passifs à pointes d'atténuation infinies réalisés avec des composants à faibles pertes. Application aux filtres de Cauer à degré élevé, *Revue Technique Thomson-Csf*, 3, n° 2 Juin 1971, p.p. 269 - 306.
5. R. Saal, E. Ulbrich, On the design of filters by synthesis, *IRE Transactions on Circuit Theory*, CT 5, December 1958, p.p. 284 - 327.
6. G. Dubé, Un ensemble de programmes relatifs aux filtres passifs, *Revue Technique Thomson-Csf*, Vol. 6, n° 1, Mars 1974, p.p. 81 - 104.
7. A. Zverev, *Handbook of Filter Synthesis*, John Wiley and Sons, New York, 1967.
8. D.S. Humpherys, *The analysis, design and synthesis of electrical filters*, Prentice-Hall, 1970.
9. G. Volluet, M. Savy, Etude d'application de l'électronique du Tantalate de Lithium, *Final Report*, November 1975, L.C.R., Thomson-Csf.
10. M. Onoe, T. Shinada, K. Itoh, S. Miyazaki, Low Frequency resonators of Lithium Tantalate, *Proc. 27th Annual Symposium on Frequency Control*, 1973.

BANDWIDTH 4 KHz

	OPERATING FREQUENCY (KHz)	QUALITY FACTOR (Q)	NORMALIZED QUALITY FACTOR (q)	TEMPERATURE COEFFICIENT PER °C	VARIATION IN Hz FOR $t = 25^{\circ}\text{C}$
L-C	25	250	40	$5 \cdot 10^{-5}$	31
MECHANICAL	100	15.000	600	$1 \cdot 10^{-6}$	2,5
CERAMIC	100	500	20	$1 \cdot 10^{-5}$	25
FERRITE	100	10.000	400	$5 \cdot 10^{-6}$	12,5
QUARTZ	5000	$1 \cdot 10^6$	80	$5 \cdot 10^{-8}$	6,3
LI TA O ₃	200	50.000	1000	$2 \cdot 10^{-6}$	10

TABLE I

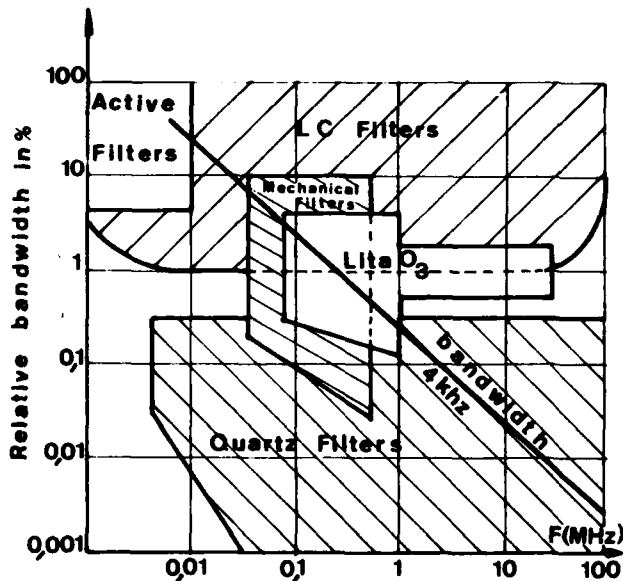


Fig. 1 - Limitations for different technologies

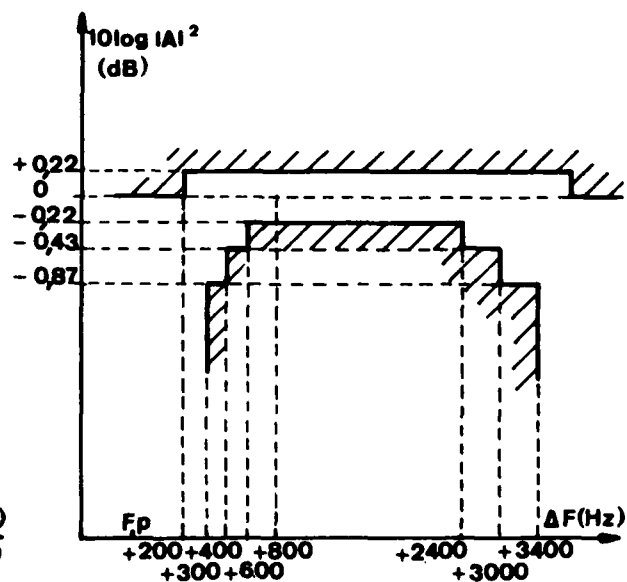


Fig. 2 - Frequency limits in the passband

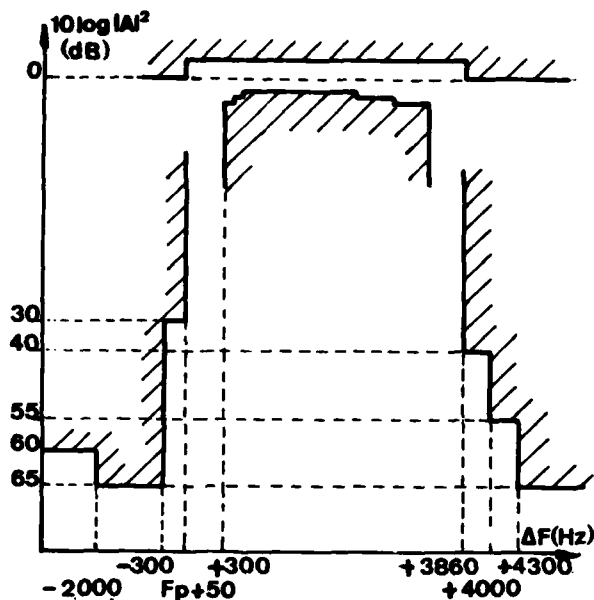


Fig. 3 - Frequency limits in the stopband

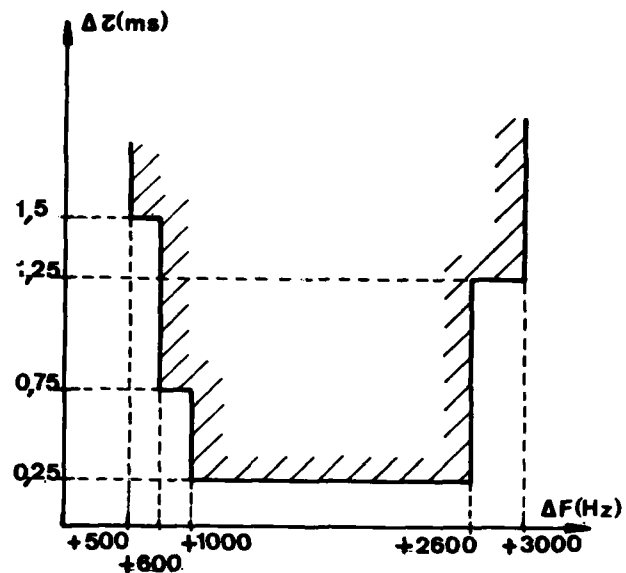


Fig. 4 - Variation of the group-delay

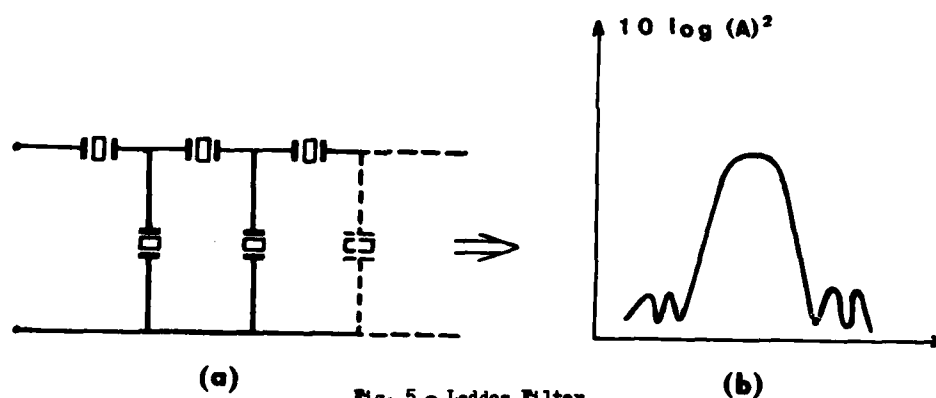


Fig. 5 - Ladder Filter

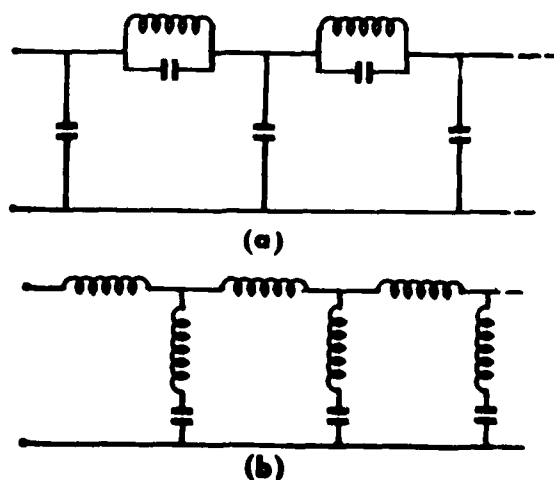


Fig. 6 - Low-pass Cauer filter after synthesis

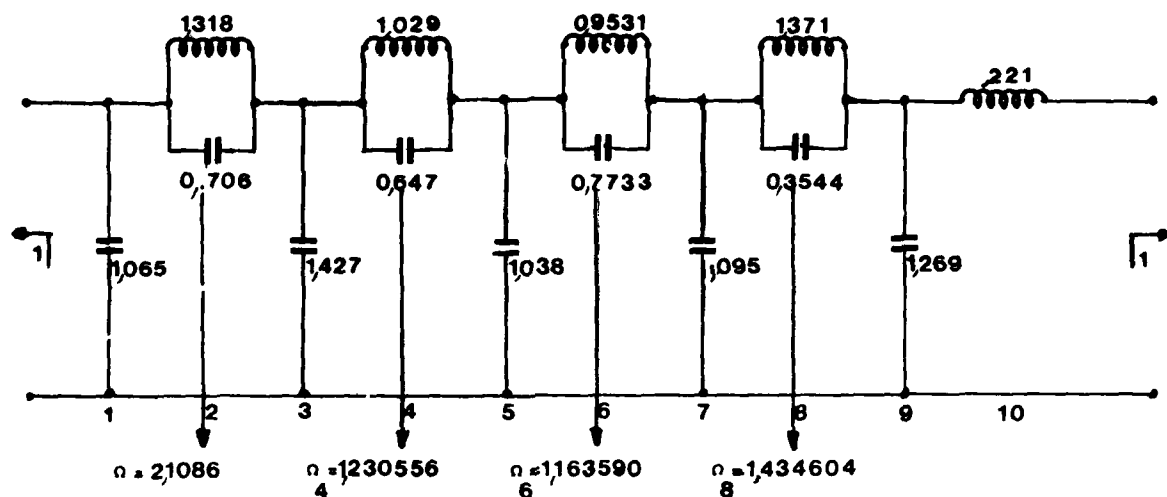


Fig. 7 - Normalized low-pass filter

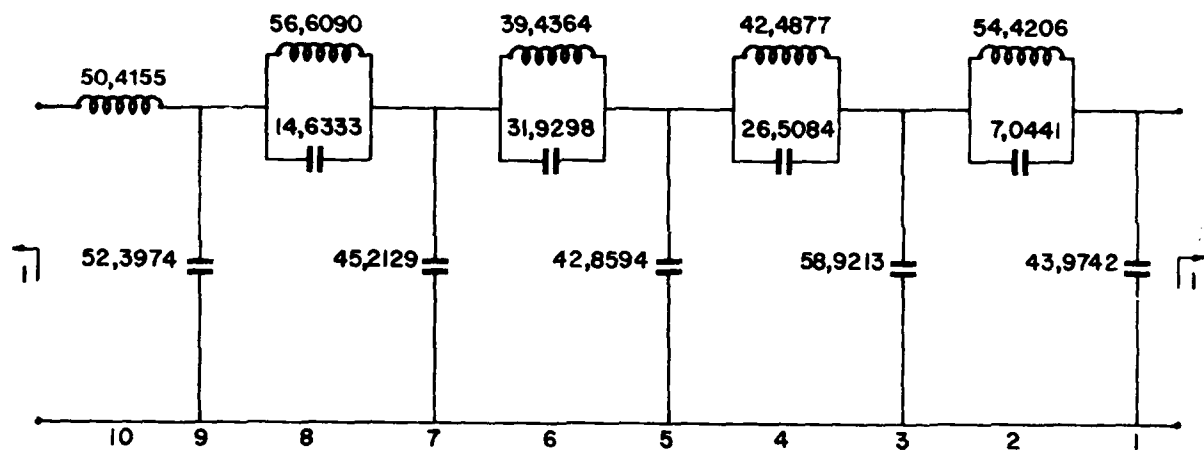
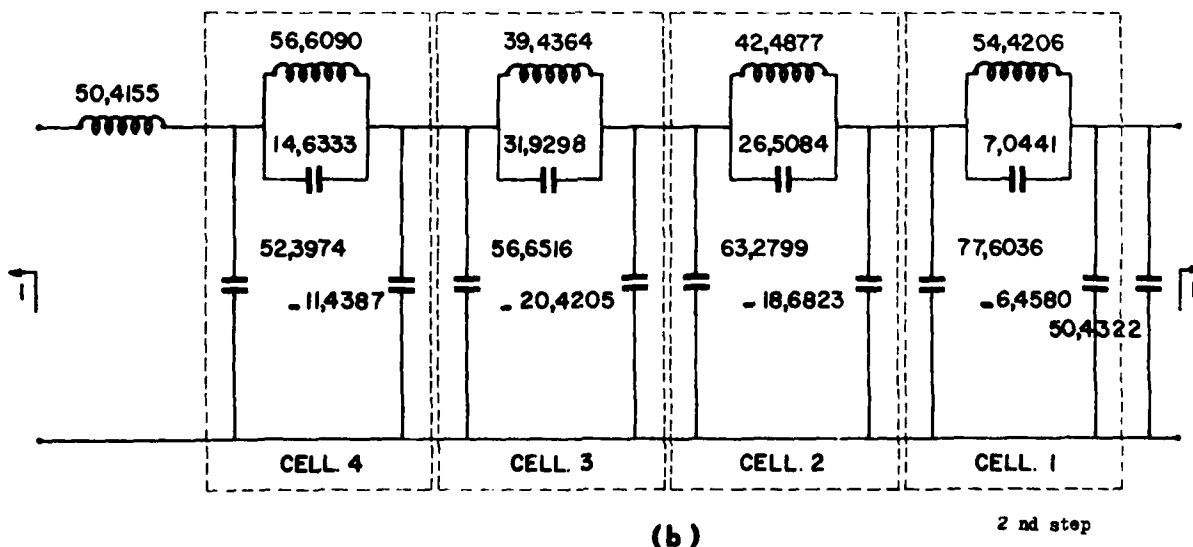


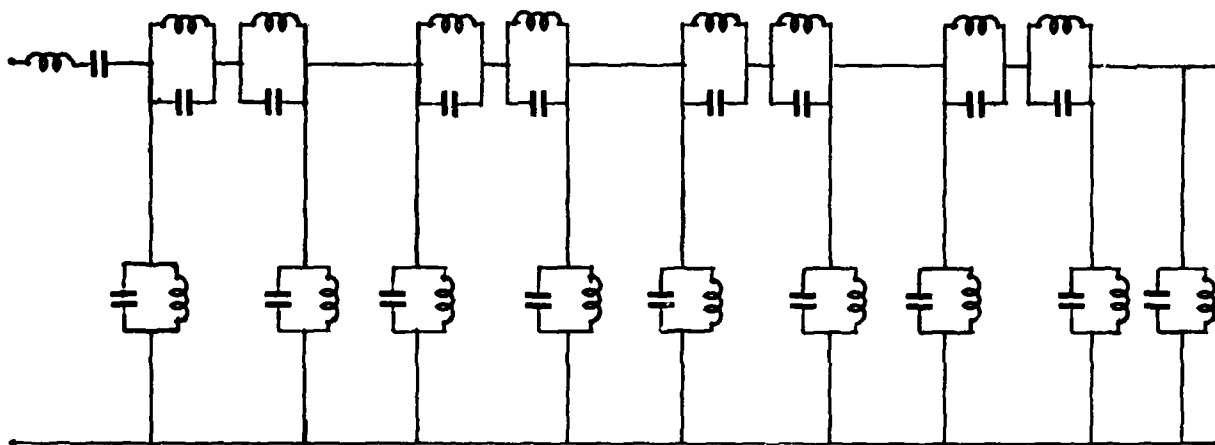
Fig. 8 - Transformation of the 1 st, 2 nd, and 3 rd steps

(a) 1 st step



(b)

2 nd step



(c) 3rd step

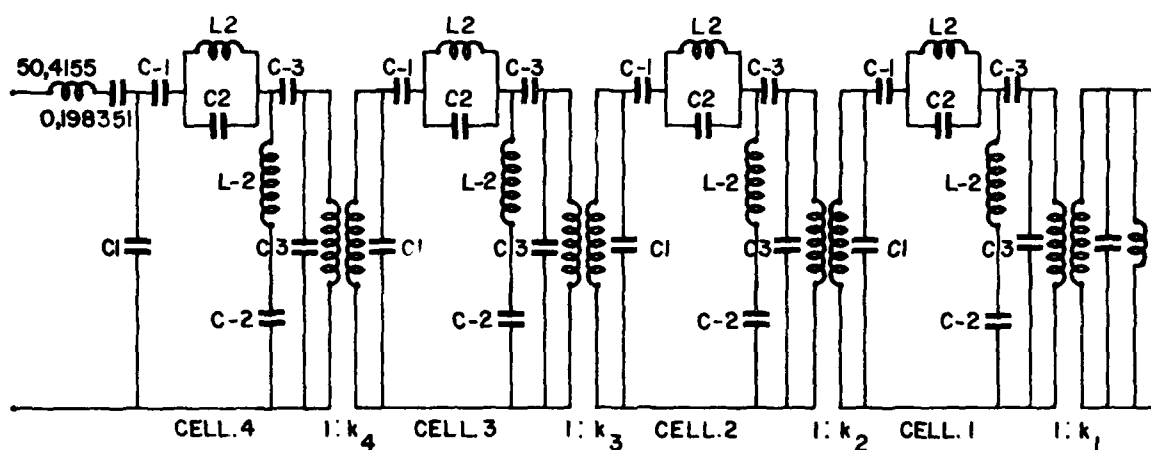
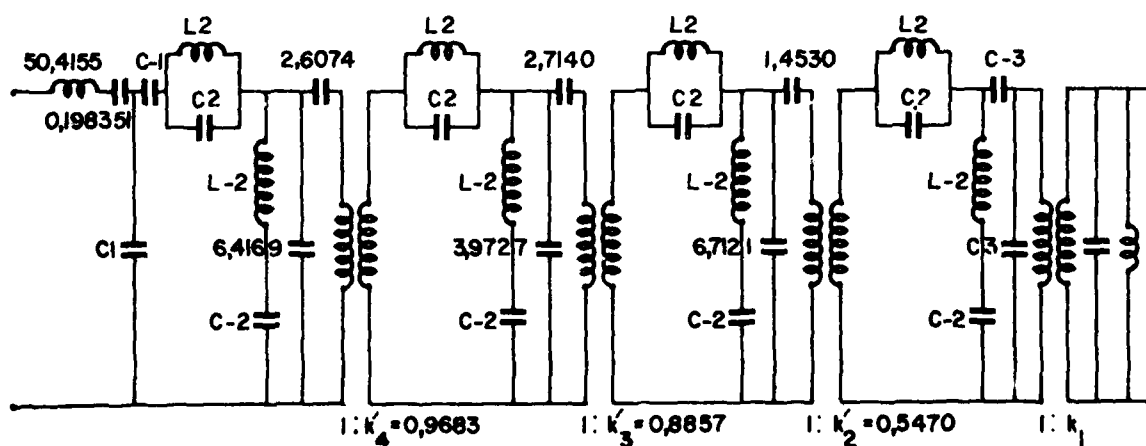
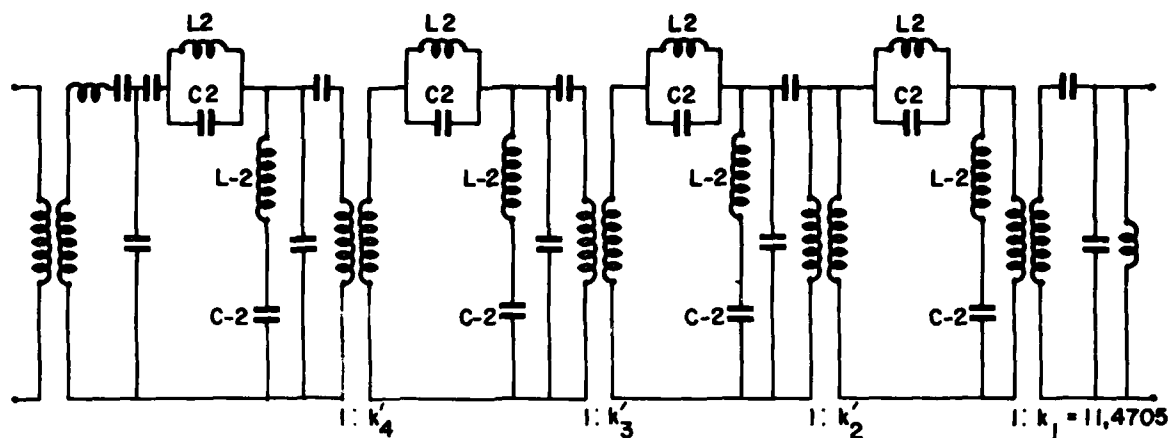


Fig. 9 • Remainder of the transformation

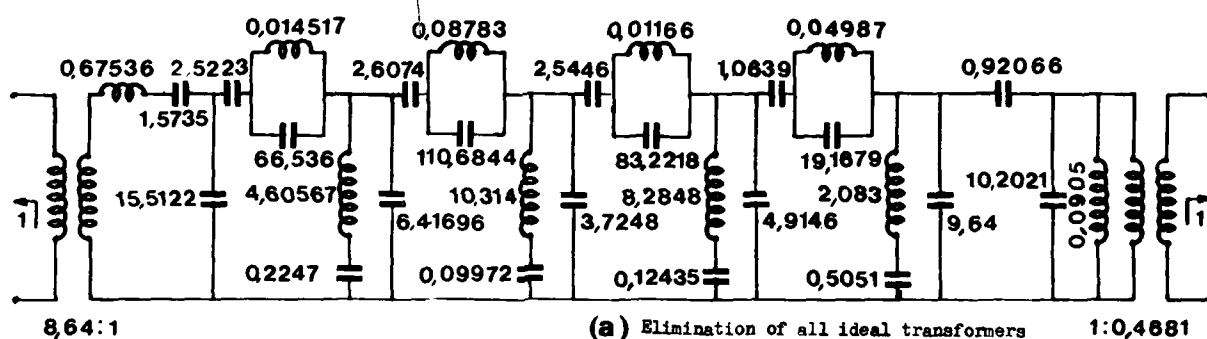
(a) 4th step



(b) 5th step

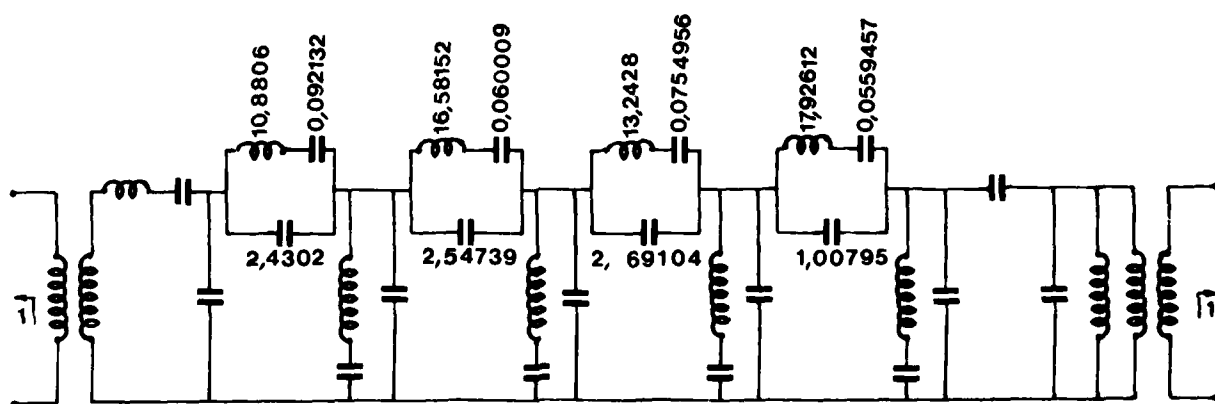


(c) 6th step



(a) Elimination of all ideal transformers

Fig. 10- End of the transformation



(b) Introduction of the piezoelectric resonators

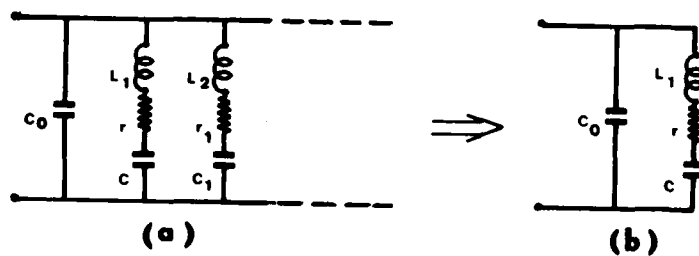
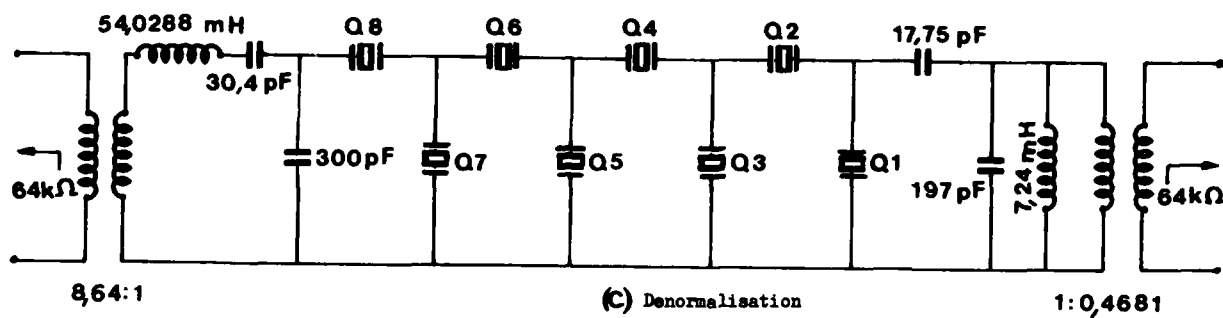


Fig. 11 -Equivalent electrical circuits

(a) - General equivalent circuit

(b) - Equivalent circuit for the fondamental

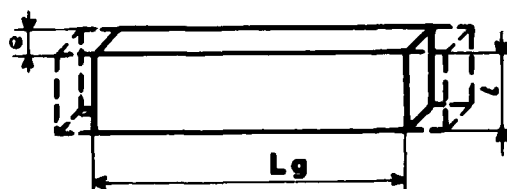


Fig. 12 -Extensional mode of vibration

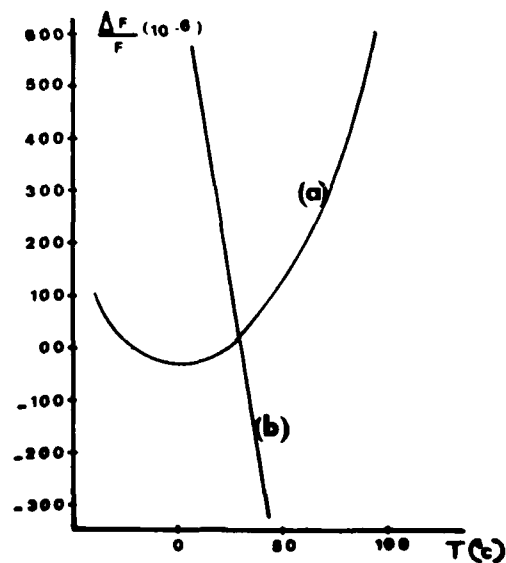


Fig. 13 -Frequency temperature characteristics

(a) Serie résonance

(b) Parallel resonance

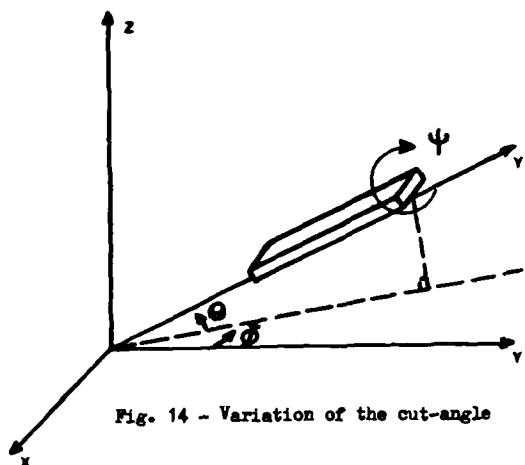


Fig. 14 - Variation of the cut-angle

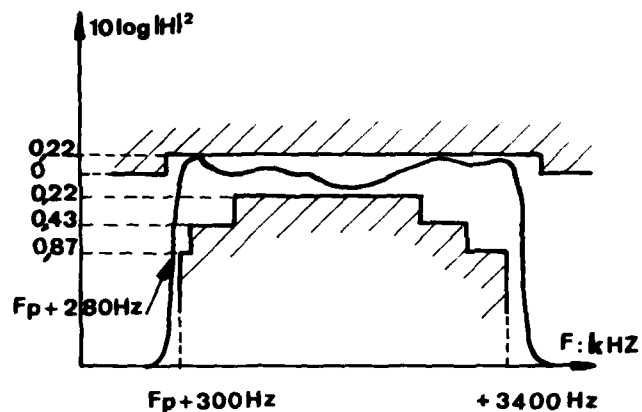


Fig. 15 - Pass-band (4 types of resonators and series inductance)

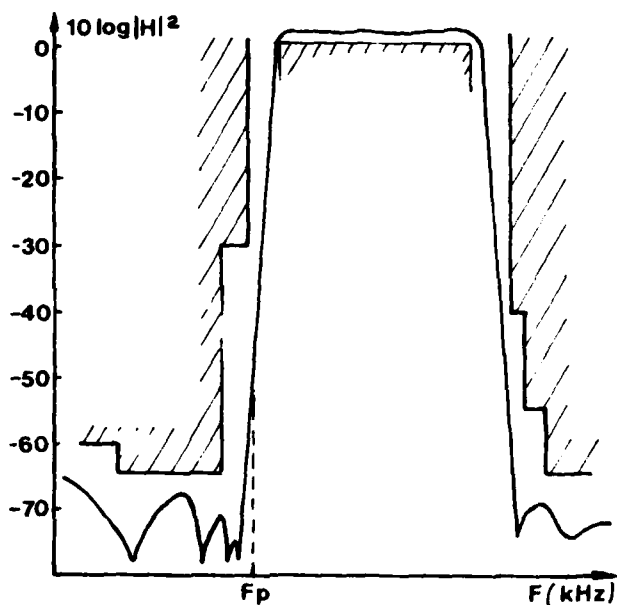


Fig. 16 - Attenuation (4 types of resonators and series inductance)

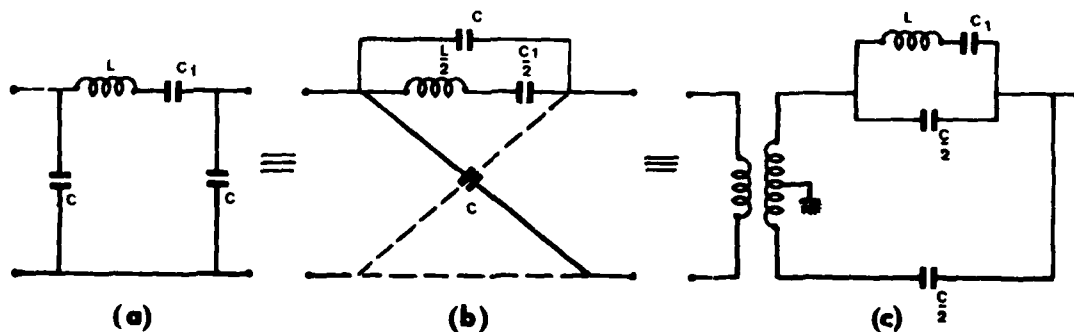


Fig. 17 - Equivalence $\pi \Rightarrow$ Jaumann

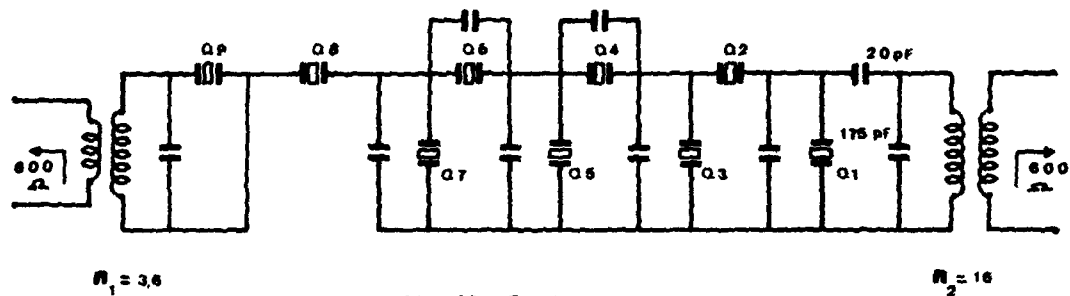


Fig. 18 - Final network

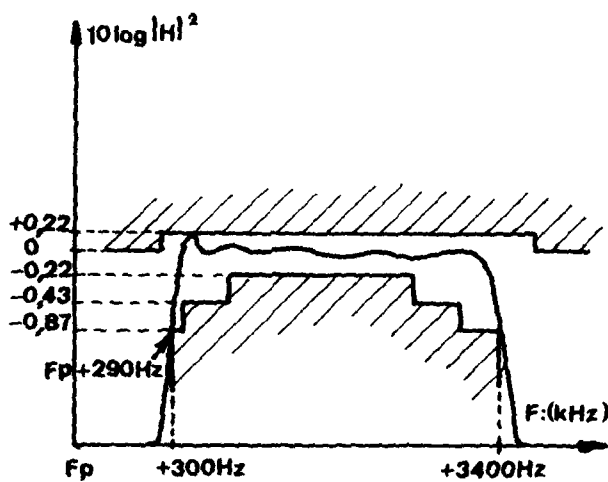


Fig. 19 - Pass-band at +25°C (9 resonators - 4 types)

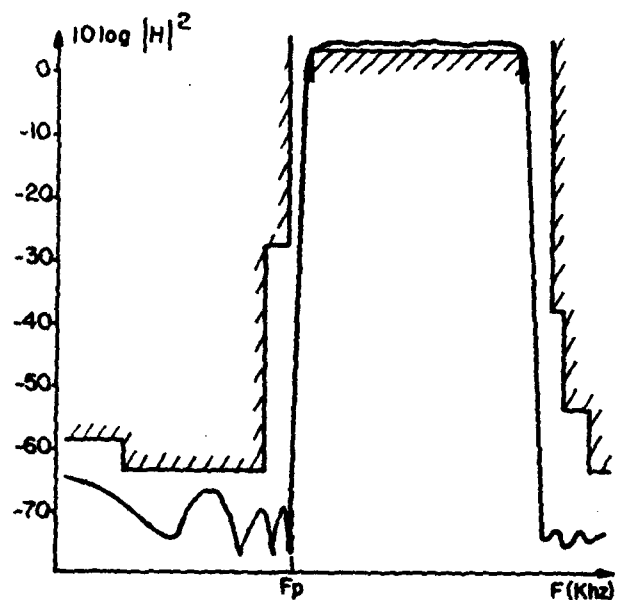


Fig. 20 - Attenuation - (9 resonators - 4 types)

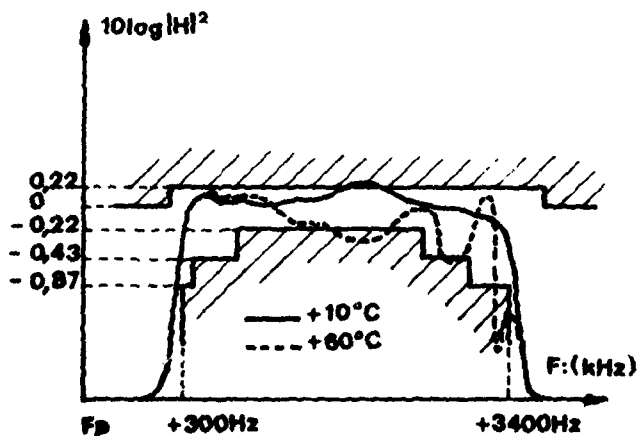


Fig. 21 - Pass-bands at +60°C and +10°C

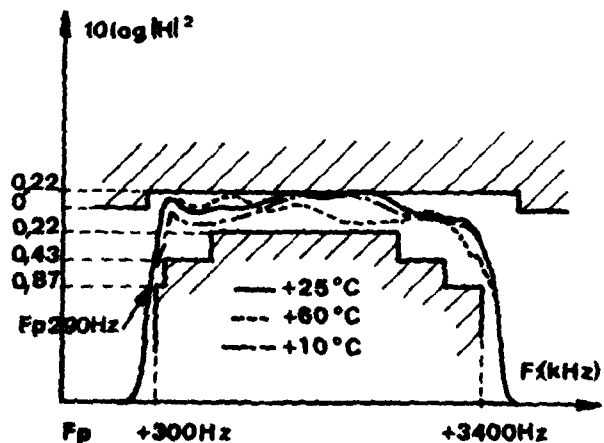


Fig. 22 - Passbands of the compensated filter

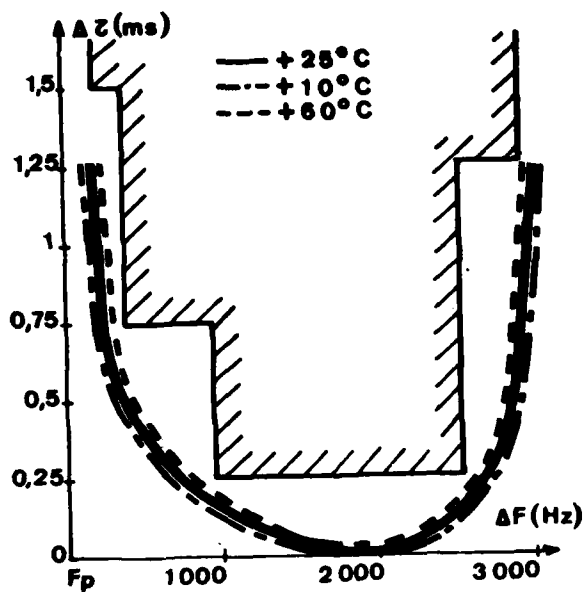
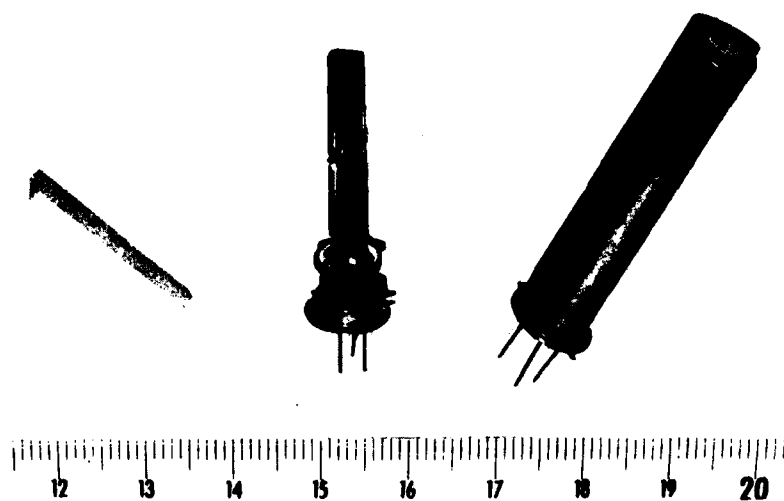
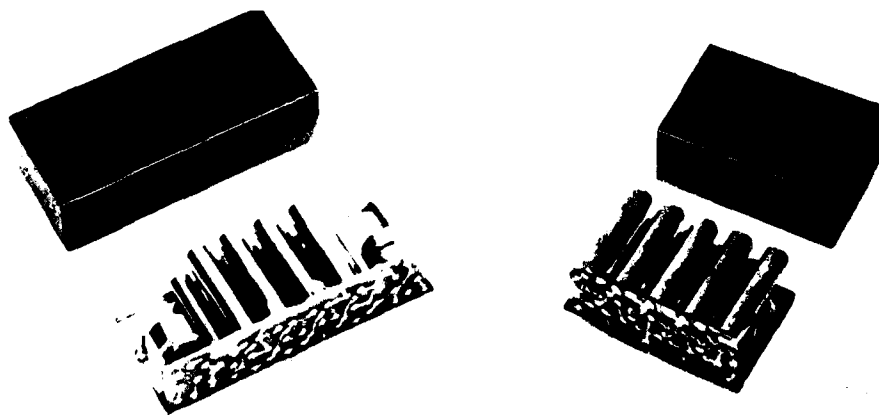


Fig. 23 - Variations of the group-delay
(compensated filter)

PHOTOGRAPH I



PHOTOGRAPH II



1.2 GHz TEMPERATURE-STABLE SAW OSCILLATOR

R.D. Weglein
Hughes Research Laboratories
3011 Malibu Canyon Road
Malibu, California 90265

Abstract

An experimental 1.2 GHz SAW delay-line oscillator is described and its performance critically analyzed. The key element of the oscillator is a 0.1% bandwidth SAW delay-line filter, consisting of a 300 Å thick Al transducer pattern containing 600 elements each 0.7 μm wide on a temperature-stable, rotated y-cut quartz platelet. The transducer thin-film pattern is fabricated by 10X reduction photolithography using projection printing from a precision reticle mask on a laser interferometer-controlled photo repeater. The parabolic frequency variation with temperature is centered near 40°C in this device; this 40°C can be positioned over a wide range in both directions by adjusting the crystal cut and metal-film thickness. The noise spectrum close to the carrier is presented and compared to prediction. Finally, the initial frequency drift characteristics of the oscillator are described and mechanisms for the variation of frequency with time are discussed.

Key words

Surface Acoustic Wave (SAW), Microwave Oscillator, Temperature Stability, Aging, Projection Lithography.

Introduction

The development of microwave SAW oscillators is currently motivated by both economic and performance considerations. First, the high-frequency operation above 1 GHz using planar, narrow-band, and temperature-stable SAW filters promises economic future procurement, in part due to the reduced need for frequency multipliers and their attendant complexity. Second, their immunity to vibration¹ permits their use in several applications where high spectral purity must be maintained in a vibration-prone environment. In this paper, the characteristics of an experimental 1.2 GHz SAW oscillator are reported. In this range, performance based on the low-frequency analysis deviates somewhat from predictions. One purpose of this paper is to describe the measured oscillator characteristics in the areas of fabrication, spectral purity, and temperature sensitivity. Some preliminary measurements on aging properties of the SAW oscillator and their relation to the low-frequency bulk crystal devices are also reported.

Basic Oscillator Design

The experimental evaluation of a SAW oscillator requires some degree of flexibility so that the effects of the individual feedback elements may be ascertained unambiguously. With this in mind, the layout of the oscillator feedback loop, shown in Figure 1, has been initially organized with discrete components that include the SAW filter, amplifier, external delay, and output coupler. A stable oscillator produces a frequency f_0 , at which the total loop phase shift ϕ_T contains an integral number (n) of 2π radians, as shown in the figure. The SAW filter design follows the mode-selection technique² with an aperture of 200 wavelengths and a mean transducer center-to-center spacing of 500 wavelengths. Each of the two transducers contains 100 electrode pairs, one in a closed array and one in a thinned-array configuration with 20 periods in each of 5 groups. Blooming sections were introduced in the

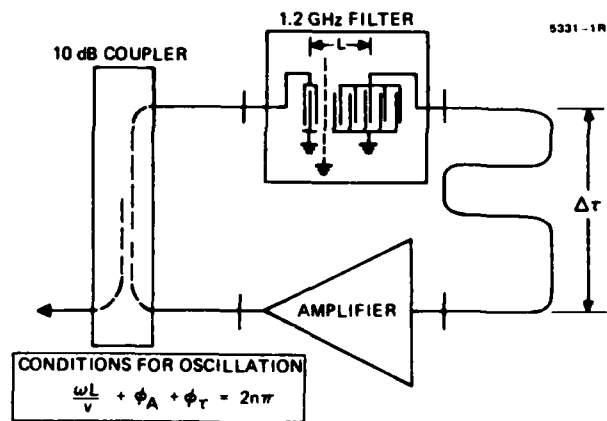


Figure 1. 1.2 GHz SAW oscillator schematic.

second transducer as a means of effectively reducing intra-transducer reflections.³ In the absence of propagation loss, this transducer configuration results in a filter group delay of 410 nsec. The resulting SAW filter was installed in a standard MIC circuit frame and connected to the 50 Ω microstrip lines with thermocompression bonded 1 mil Au wires. The package was designed in such a way that the bonding wire length provides the necessary inductive tuning for series resonance of 1.2 GHz. In addition to the SAW filter, the experimental oscillator assembly, shown in Figure 2, includes a hybrid production amplifier (gain = 23 dB, saturated output power = +13 dBm), a rigid cable section for external delay, and a commercial 10 dB coupler. The large decoupling simultaneously minimizes loop loss contributions, which raise the oscillator noise floor, and provides for increased frequency stability in the face of impedance mismatch variations.

SAW Filter

The SAW filter patterns, described above, were fabricated on a large quartz plate of 38° rotated, Y-cut crystal orientation. Projection printing⁴ from a 10X reticle produced up to 10 patterns on 5 mm x 10 mm centers (see Figure 3) by the step-and-repeat method with lift-off lithography. The Al metallization provides film patterns 300 to 350 Å thick each containing 600 electrodes approximately 0.7 μm wide and 526 μm long. A portion of one such pattern is shown in Figure 4. The electrode width, slightly wider than the desired 0.66 μm, has a bearing on the resulting filter center frequency and on the accuracy with which the turnaround temperature T_0 can be predicted. The latter aspect is treated in a later section of this paper.

The frequency response of the series-tuned SAW filters is shown in Figure 5. The frequency at the minimum insertion loss of 20 dB corresponds to the oscillator frequency of 1.187 GHz, slightly more than 1% lower than the design frequency of 1.2 GHz. By far the largest contributor to this error is the velocity

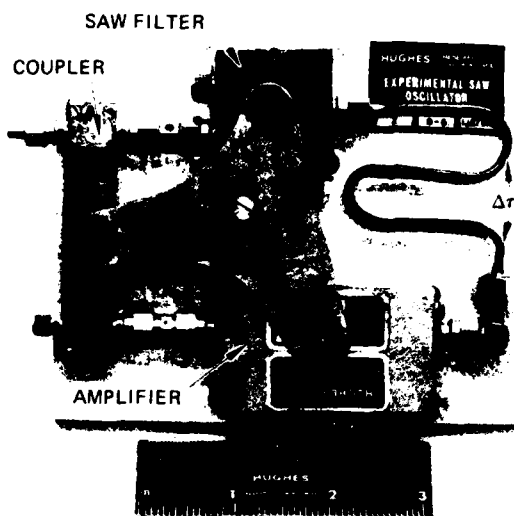


Figure 2. Experimental 1.2 GHz SAW oscillator assembly.

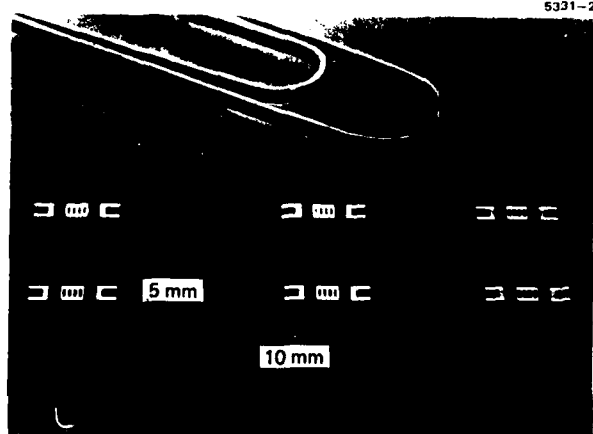


Figure 3. Multiple SAW filter patterns on quartz substrate by 10X projection lithography.

slowing caused by the metallic grating structures; this velocity slowing is several times larger⁵ than that predicted from perturbation theory.⁶ The measured insertion loss is consistent with a propagation loss of 6.5 dB/μsec at this frequency. Further contributions to the insertion loss result from a thin-film ohmic transducer loss that corresponds to about three times the bulk resistivity for Al and to the distortion⁷ in the closed array transducer response produced by mass loading. The predicted transducer responses do not coincide. Their relative shift is due to the large velocity perturbation that was also responsible for the reduced frequency, discussed above, producing a lesser frequency shift in the thinned array transducer. Some evidence of triple transit echo is seen in the rippled slope of the central pass band. The largest secondary peak near 1199 MHz (-11 dB relative to the central response) is sufficiently

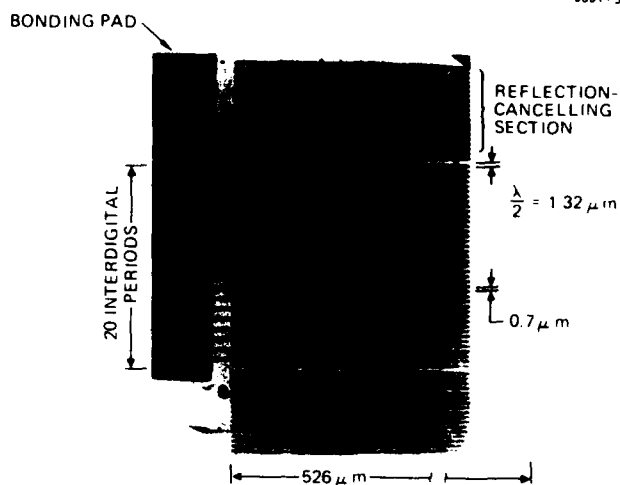


Figure 4. Portion of interdigital transducer pattern.

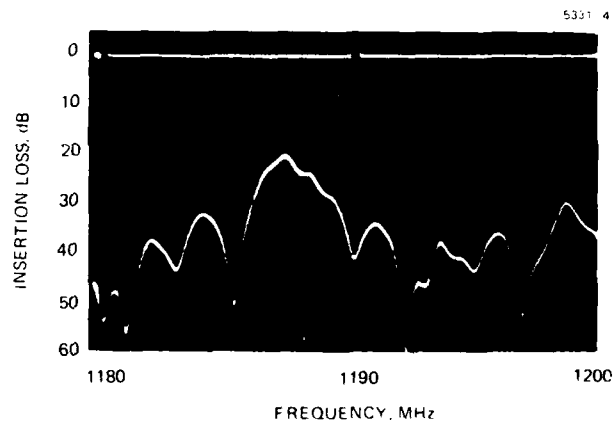


Figure 5. Frequency response of 1.2 GHz SAW filter.

low to preclude measurable contribution in the oscillator noise spectrum; it is also consistent with predictions.⁸

Oscillator Noise Spectrum

The noise characteristics of the SAW delay line oscillator have been amply treated in the literature.³ The noise spectral density may be defined by two relevant parameters, the noise floor

$$N_f = \frac{A k T F}{P_{\text{loop}}} \quad (1)$$

and the intercept frequency

$$f_N = (2\pi \tau_g)^{-1} \quad (2)$$

below which the noise spectrum rises at a 20 dB/decade rate. In these expressions, A and F are the amplifier power gain and noise figure, respectively, k is the Boltzman constant, T the ambient temperature, and τ_g is the group delay of the SAW filter. The group delay for a particular filter may be obtained from its relationship to the phase slope near the filter center frequency, given by

$$\frac{d\phi}{df} = 2\pi f \tau_g \quad (3)$$

A network analyzer and a variable frequency or sweep-signal source may be used to determine $d\phi/df$.

Using the parameters shown in Table I, the predicted noise intercepts and the experimental noise spectral density are plotted in Figure 6. The experimental data was measured by two different methods. The solid curve was obtained at X-band with an 8-times multiplier on a standard Hughes noise test set. The test set involves a microwave discriminator using a stable high-Q cavity resonant near 9.6 GHz. The resulting data was normalized to 1.2 GHz by subtracting 18 dB (or 20 log 8), where zero noise contribution from the frequency multiplier was presumed. This curve agrees well with predictions except near the noise floor, where the sensitivity of the noise test set itself is measured rather than the oscillator noise floor.

Table I. Feedback Loop Parameters

Amplifier gain	23 dB
Noise figure	4 dB
Loop power	+10 dBm
Group delay	0.397 sec

The second measurement method involves two 1.2 GHz SAW oscillators with approximately equal parameters. One oscillator is phased locked (loop bandwidth = 100 Hz) to the second one by introducing a voltage-sensitive phase-delay element in its feedback path. The two oscillators are therefore at the same frequency but in phase quadrature. The resulting noise spectrum is measured on a low-frequency spectrum analyzer, and the results (dashed curve) are plotted under the assumption that both sources contribute equally. Spectral noise data obtained from this type of measurement may be in error at low offset frequencies unless adequate isolation exists between the two oscillators and unless the bandwidth of the phase-locked loop is smaller than the lowest offset frequency noise component to be measured. Thus, at offset frequencies less than 200 Hz, the phase-locked loop noise data tends to bend erroneously over and cross the predicted 20 dB/decade line. The loop bandwidth is currently being reduced to 10 Hz to improve the accuracy of low offset frequency measurements.

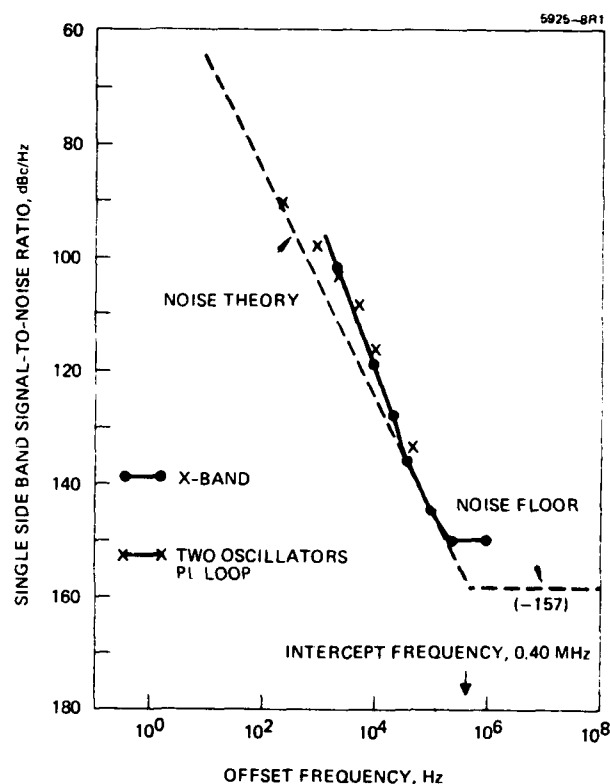


Figure 6. 1.2 GHz SAW oscillator noise spectrum.

Turnaround Temperature

The zero temperature coefficient of delay (TCD) for surface acoustic waves on the free surface of X-propagation quartz is a function of the rotated Y-cut angle θ . In addition, the presence of a continuous Al film on the quartz SAW substrate has been shown to reduce the turnaround temperature T_0 at which the zero TCD occurs.¹⁰ Recently, the effect of a periodic metallic transducer grating on quartz has been theoretically calculated and experimentally verified for the case of continuous grating of equal stripe-to-gap ratio.¹¹ These results have been used to compute the temperature shift ΔT_0 as a function of the normalized film thickness $2\pi h/\lambda$ and fractional metal coverage δ , typical of SAW oscillator filter patterns, under the assumption that T_0 varies linearly with δ . The linear dependence assumption is based on the work reported in Ref. 10. It is further assumed that this function is independent of the plate-normal angle θ over the range $32^\circ < \theta < 43^\circ$, where $\theta = 42.75^\circ$ corresponds to the popular ST-cut orientation. Computations are shown in Figure 7 for an Al film on X-propagating quartz. For typical metal coverage found in SAW oscillator filter patterns, the reduction in T_0 may be substantial at microwave frequencies, since the Al film cannot reasonably be made thinner. These curves and those from Ref. 9 were used to design the 1.2 GHz oscillator with a T_0 slightly above room temperature. If T_0 is to be at 30°C , for example, then, with the desired film thickness ($h = 300 \text{ \AA}$) and metal coverage ($\delta = 0.24$), a downward shift in T_0 of 26°C , as predicted from Figure 7, would require a rotated Y-cut angle $\theta = 38^\circ$.

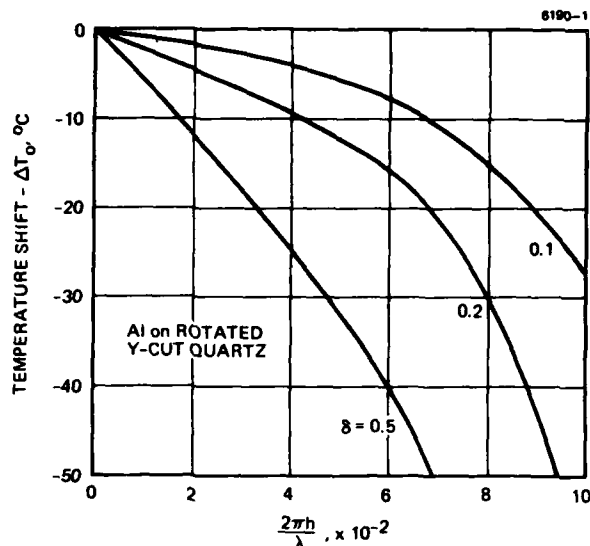


Figure 7. Effect of metal coverage on turnaround temperature T_0 .

The frequency variation with temperature of two SAW oscillators and one SAW filter are shown in Figure 8, all of which have the 38° plate normal orientation. Although the SAW devices used in the oscillators were of identical design, the SAW filters did not use the bloomed anti-reflection gratings; this reduced the metal coverage. The T_0 's of the two oscillators are located at approximately 18°C and 31°C ; the T_0 of the SAW filter is at 40°C . The results from these and from several oscillators of somewhat lower frequencies are summarized in Table II. Here, T_0 refers to the measured turnaround point, T_{of} to that of the free surface, and ΔT_{exp} and T_{th} denote the measured and predicted temperature shifts, respectively.

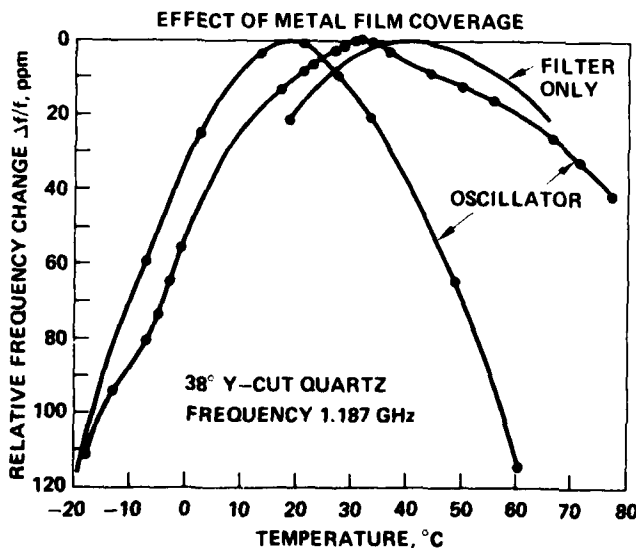


Figure 8. SAW oscillator temperature characteristic.

Table II. Turnaround Temperature versus Metal Coverage⁽¹⁾

Frequency, MHz	$2\pi h/\lambda$	δ	T_0 , °C	T_{of} , °C	$-\Delta T_{exp}$, °C	$-\Delta T_{th}$, °C
1200 ⁽²⁾	0.05	0.16(0.24)	40	56	16	9 (17)
1200	0.072	0.24(0.36)	33	56	23	26 (40)
1200	0.072	0.24(0.36)	20	56	36	26 (40)
1005 ⁽³⁾	0.06	0.20(0.29)	33	56	23	16 (22)
1005	0.06	0.20(0.29)	40	56	16	16 (22)
1005	0.06	0.20(0.29)	35	56	19	16 (22)

⁽¹⁾ 38° rotated Y-cut quartz

⁽²⁾ Stripe to period ratio = 0.7

⁽³⁾ Stripe to period ratio = 0.6

⁽⁴⁾ () assumes transducer center-to-center spacing.

These results show that the turnaround points may be approximately predicted, but that some scatter should be expected. The difference in T_0 between the two oscillators probably arises from variations in the stripe-to-gap ratio, which is difficult to measure accurately in these fine geometries.

Frequency Drift with Time

The aging characteristics of SAW oscillators have received relatively little attention.^{12,13} Substantial similarity appears to exist between SAW and bulk crystal devices insofar as aging mechanisms are concerned. In part, this similarity arises from the materials common to both devices, namely, the quartz crystal plate and the Al thin-film electrodes. Important aging mechanisms relevant to both SAW and bulk devices are given in Table III.

Table III. Acoustic Oscillator Aging Mechanisms

Mechanism	Activation Energy, eV
Metal film surface oxidation	0.21
Metal film stress relief	-
Metal film substrate adhesion	-
Diffusion across interface	<3
Substrate defect stress relief	-
Contamination	-
Encapsulation-related stress	-
Loss of hermeticity	-

For the most part, the aging rates are not known. Our observations with the 1.2 GHz SAW oscillator supports the premises that frequency drift resulting from surface oxidation on the Al film is quite rapid (low activation energy) and that the initially observed negative aging rate during the "burn-in" period is directly related to the mass increase caused by oxidation.

The initial aging rate (frequency drift) of one of our 1.2 GHz SAW oscillator was monitored continuously for more than 80 hr using a frequency-counter/D/A-converter/chart-recorder combination (see

Figure 9). Immediately after fabrication, the SAW device used in this test was assembled (see Figure 2) without the special post-processing commonly followed in the assembly of bulk crystal resonators. The aging rate was approximately exponential with a logarithmic decrement of about 9 hr, which is in substantial agreement with the known oxidation rate for Al_2O_3 film growth.¹⁵ The total frequency drift during the observed period was 35 ppm.

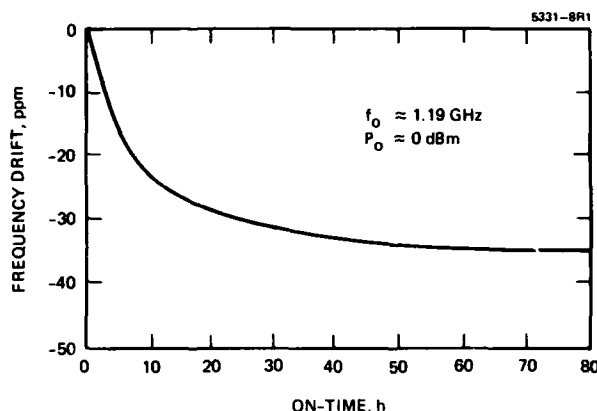


Figure 9. "Burn in" characteristic of 1.2 GHz SAW oscillator.

The effect of a mass increase on frequency change resulting from surface oxidation may be estimated from perturbation theory,⁶ as shown in Figure 10. The estimation is made under the assumption that the observed frequency shift (velocity slowing) is due to the increased average mass of the oxidized film. The mean density ρ of the oxidized film is larger than the Al film of thickness h ($\rho = 2.7 \text{ gm/cm}^3$) because of the higher density of the Al_2O_3 surface layer of thickness Δh ($\rho = 3.96 \text{ gm/cm}^3$). If one assumes that the velocity slowing varies as $\rho^{0.5}$, the expected fractional frequency change is given by

$$\frac{\Delta f}{f_0} \approx 4.6 \times 10^{-4} \frac{\Delta h}{h} \quad (4)$$

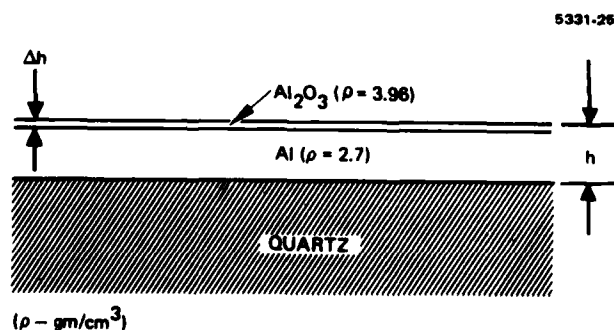


Figure 10. Oxidized Al-film on quartz substrate.

Therefore, a 10% change in oxidized film thickness ($\approx 30 \text{ Å}$) would lead to a downward frequency drift of 46 ppm. The observed frequency drift during the 80 hr period has the correct sign and falls well within the range of both the magnitude and formation rate known for the oxidation of Al films exposed to

ambient atmosphere.¹⁵ The low activation energy of this process ($E_a \approx 0.21 \text{ eV}$) precludes long-term effects (i.e., longer than about 30 days) at room temperature. Several pre-aging processing techniques practiced in bulk crystal technology may be used to reduce this aging mechanism (and others listed in Table III) and thereby increase long-term stability in high-frequency SAW oscillators.¹⁶

Summary and Conclusions

The characteristics of an experimental 1.2 GHz SAW oscillator, described in this paper, are summarized in Table IV. Modular construction was used so that the individual components of the feedback loop could be separately evaluated and easily interchanged. Projection lithography (10X) yields the excellent definition in 0.7 μm electrode geometry that is required for this frequency range. In spite of the larger than anticipated effects of the metal film on frequency, the result of velocity slowing, a 20 dB device insertion loss was obtained with a 3 dB bandwidth of 0.08%. The insertion loss of the largest side lobe in the frequency response is 11 dB higher. Good agreement is obtained between the measured noise spectrum and the theoretical prediction; the prediction is characterized by an intercept point offset 0.4 MHz from the oscillator frequency and a frequency-independent noise floor of -157 dBm.

Table IV. Characteristics of the 1.2 GHz SAW Oscillator⁽¹⁾

Frequency	1.187 GHz
Circulating power	+10 dBm
Power output (10 dB coupler)	+0.3 dBm
Temperature sensitivity at 26°C	$8 \times 10^{-7}/^\circ\text{C}$
Temperature sensitivity	-31×10^{-9} $(T - 30)^2 (^\circ\text{C})^2$
Frequency stability	$< 10^{-9}$
$10^{-5} < \tau < 10 \text{ sec}$	
Filter insertion loss	20 dB
Single-sideband noise power per Hz at 0.25 MHz	-148 dBc (-157 dBc)
(1) 38° rotated Y-cut quartz	

The effects of mass loading on T_0 , the zero temperature coefficient turn-around temperature, were discussed. Temperature stability near 30°C was achieved by properly choosing the rotated Y-cut angle ($\theta = 38^\circ$) so as to offset the downward shift in T_0 due to the metal film. Finally, one of the aging mechanisms, Al-film surface oxidation was observed and its effect on long term stability analyzed.

During this study, no fundamental problems were encountered that would preclude the extension to higher microwave frequencies. The use of the space harmonic transducers with the 0.7 μm electrode geometry could readily extend the frequency to 1.8 GHz, including a moderate decrease in the effective oscillator Q. Further increases in the frequency of microwave SAW oscillators are possible through high-resolution electron-beam fabrication¹⁷ and the employment of a new mode of propagation using surface-skimming bulk waves.¹⁸

Acknowledgments

Several co-workers contributed their expertise to this program. The efforts of A. R. Janus and L. Dyal are gratefully acknowledged for the SAW pattern fabrication. B. E. Dobratz and A. Miller supplied the noise spectral data. R. Dimon and R. Maheux aided in the assembly and measurement of the SAW device and oscillator. Particular thanks are extended to O. W. Otto for helpful discussion and critical review of the manuscript.

References

1. R. D. Weglein and O. W. Otto, "The Effect of Vibration on the Noise Spectra of SAW Oscillators," *Electron. Lett.*, Vol. 13, No. 4, 103-104 (17 Feb 77).
2. J. Crabb, M. F. Lewis, and J. D. Maines, "SAW Oscillators: Mode Selection and Frequency Selection," *Electron. Lett.*, Vol. 9, No. 10, 195-197 (17 May 1973).
3. M. F. Lewis, "The SAW Oscillator—A Natural and Timely Development of the Quartz Crystal Oscillator," *Proceedings 28th Annual Frequency Control Symposium (E.I.A., 2001 Eye St., NW, Washington, DC 20006)*, 304-314 (1974).
4. A. R. Janus, L. Dyal, and H. M. Gerard, "Projection Printing for the Fabrication of L-Band SAW Transducers," *Proceedings 1975 Ultrasonic Symposium, IEEE Cat. No. 75 CH0 994-4SU*.
5. A. J. Slobodnik, Jr., et al., "SAW Filters for Frequency Synthesis Application," *Proceedings 1976 Ultrasonic Symposium, IEEE Cat. No. 76 CH 1120-5SU*.
6. D. Penunuri and K. M. Lakin, "Surface Acoustic Wave Velocity of Isotropic Metal Film on Selected Cuts of $\text{Bi}_{12}\text{GeO}_{20}$, Al_2O_3 and LiNbO_3 ," *IEEE Trans. Sonics and Ultrasonics* **21**, 293-295 (Oct 1974).
7. T. W. Bristol, W. R. Jones, P. B. Snow, and W. R. Smith, "Application of Double Electrodes in Acoustic Surface Wave Design," *Proceedings 1972 Ultrasonic Symposium, 343-345, IEEE Cat. No. 72 CH0 708-8SU*.
8. O. W. Otto and R. D. Weglein, "SAW Oscillators Using Reflective Gratings," *Proceedings 1975 Ultrasonic Symposium, 225-260, IEEE Cat. No. 75 CH0 994-4SU*.
9. J. F. Dias, et al., "The Temperature Coefficient of Delay Time for X-Propagating Acoustic Surface Waves of α -Quartz," *IEEE Trans. Sonics and Ultrasonics*, 46-50 (Jan 1975).
10. S. J. Kerpel, "Design of Harmonic SAW Oscillator Without External Filtering and New Data on the Temperature Coefficient of Quartz," *Proceedings 1974 Ultrasonic Symposium, 276-281, IEEE Cat. No. 74 CH0 896-1SU*.
11. Y. Shimizu, et al., "Temperature Dependence of SAW Velocity for Metal Film on α -Quartz," *Proceedings 1976 Ultrasonic Symposium, 519-522, IEEE Cat. No. 76 CH 1120-5SU*.
12. S. C. Gratze and R. K. Barton, "Surface Acoustic Wave Oscillator—Long Term Stability," *Electronic Engineering*, 49-51 (Mar 1975).
13. D. T. Bell, "Aging Effects in Plasma Etched SAW Resonators," *Proceedings 30th Annual Frequency Control Symposium (E.I.A., 2001 Eye St., NW, Washington DC 20006)*, 358-365 (1976).
14. A. Gerber and R. A. Sykes, "State of the Art—Quartz Crystal Units and Oscillators," Chap. 2, Part A—"Time and Frequency: Theory and Fundamentals," *National Bureau of Standards Monograph 140* (May 1974).
15. V. E. Bottom, "A Novel Method of Adjusting the Frequency of Aluminum Plated Quartz Crystal Resonators," *Proceedings 30th Annual Frequency Control Symposium (E.I.A., 2001 Eye St., NW, Washington, DC 20006)*, 96-108 (1976).
16. J. R. Vig, et al., "Surface Studies for Quartz Resonator," *Proceedings 28th Annual Frequency Control Symposium (E.I.A., 2001 Eye St., NW, Washington, DC 20006)*, 96-108 (1974).
17. E. D. Wolf, F. Ozdemir, and R. D. Weglein, "Precision Electron Beam Micro Fabrication of Acoustic Surface Wave Devices," *Proceedings Ultrasonic Symposium, 510-516 (Nov 1973), IEEE Cat. No. 73 CH0 807-8SU*.
18. T. I. Browning and M. F. Lewis, "New Family of Bulk-Acoustic-Wave Devices Employing Interdigital Transducers," *Electron. Lett.*, Vol. 13, No. 5, 128-130 (3 Mar 77), and paper (Session A) at this symposium.

TUNING QUARTZ SAW RESONATORS BY OPENING SHORTED REFLECTORS

Ronald C. Rosenfeld, Thomas F. O'Shea, & Steven H. Arneson
Motorola, Inc. Communications Division
Schaumburg, Illinois 60196

ABSTRACT. High Q surface-acoustic-wave (SAW) resonators, like bulk-wave resonators, must be tuned to precise resonant frequencies. At VHF and UHF the required tuning accuracy may be a few parts per million. One technique for tuning SAW resonators is to fabricate the resonators with reflector strips that are shorted at both ends and then to open the reflectors with a laser until the desired frequency is reached. The frequency shift is caused by the surface-wave velocity difference under shorted and floating reflectors. In this paper, experimental data is presented for the tuning of single-pole, 155 MHz, aluminum on ST-quartz resonators and a concise theoretical expression is developed that accurately models the frequency shift. The effects of tuning on transverse modes for a single and two pole resonator are discussed as well as the effects of tuning on the series resistance of a single pole resonator.

Introduction

Resonator Geometry

Single-pole surface-acoustic-wave (SAW) resonators consist of two sets of periodic reflector arrays that form a resonant cavity with a thin-film interdigital transducer that couples energy to the cavity as shown in Figure 1. The reflectors are commonly formed by grooves in the substrate or aluminum strips on the substrate. Grooved-reflector resonators have demonstrated the highest Q while aluminum-reflector resonators are simpler to fabricate since the reflectors and transducer are fabricated in the same processing steps. Also, metalized-reflector resonators are easily tuned by the technique discussed here.

RESONATOR TUNING BY OPENING SHORTED REFLECTORS

RESONATOR GEOMETRY:



FIGURE 1

EQUIVALENT CIRCUIT:

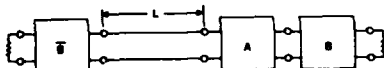


FIGURE 2

Tuning

If aluminum reflectors are shorted at the ends, as shown in Figure 1, the resonant frequency can be easily adjusted by opening the reflectors with a scribe or laser. The frequency shifts upward because the surface wave velocity is higher under the opened reflectors than under the shorted reflectors.

Marshall¹ demonstrated a 300 ppm shift between completely shorted and completely opened reflectors on ST-quartz substrates. In the following discussion, the frequency shift will be examined for partially opened reflector arrays and a useful theoretical model will be derived.

Theoretical Analysis

Model

The frequency shift is calculated by modeling the opened and shorted reflector arrays as two-port junctions using scattering parameters, joining the junctions to form a resonator circuit, and then deriving the frequency shift as a function of geometry, acoustic impedance mismatch, velocity change and number of opened and shorted reflectors. The resonator circuit is shown in Figure 2 where junctions A and B represent the opened and shorted sections of one array, B represents the other reflector array, and the cavity is represented by a transmission line of length L. The scattering coefficients for each reflector array are calculated from Sittig and Coquin². Approximate expressions for magnitude and phase of reflections from the combined reflector array are derived which are valid under the constraints that acoustic impedance mismatch, opened and shorted velocity difference, and frequency shifts are small. The effect of a transducer in the cavity can be ignored since the transducer characteristics do not change significantly over small frequency shifts.

Cavity Resonant Frequency

In resonators, resonance occurs when the round-trip wave propagation phase adds constructively. Thus, for the cavity in Figure 2, the resonant phase condition is

$$-2kL + \phi + \bar{\phi} = -2\pi m \quad (1)$$

where ϕ and $\bar{\phi}$ are the phase of the reflection coefficients, $\Gamma = \Gamma_0 e^{j\phi}$, $\bar{\Gamma} = \bar{\Gamma}_0 e^{j\bar{\phi}}$, $k = 2\pi f/v_0$, and $L = n\lambda_0/2$. The "barred" terms refer to the left reflector array. Throughout this analysis it will be assumed that resonance of a cavity formed by completely shorted reflectors occurs

at the center frequency f_0 of the reflector arrays with wavelength λ_0 , that the reflection phase ϕ of a shorted array is zero (or a multiple of 2π) at resonance, and that the velocity v_0 in the cavity and shorted reflectors are equal so that $n = m$ in the above equation. These assumptions affect only the absolute resonant frequency, not the frequency shift to be calculated here.

It will be shown later that the reflection coefficient phases can be expressed as

$$\begin{aligned}\phi(f) &= \phi_1 \Delta f / f_0 + \phi_2 \Delta v / v_0 \\ \bar{\phi}(f) &= \bar{\phi}_1 \Delta f / f_0 + \bar{\phi}_2 \Delta v / v_0\end{aligned}\quad (2)$$

where $\Delta f = f - f_0$, $\Delta v = v - v_0$, with f being the new resonant frequency and v_0 the velocity under the opened reflectors. ϕ_1 and ϕ_2 are functions of the reflector array geometry and acoustic impedance mismatch. Substituting Equation (2) into Equation (1) leads to an explicit expression for resonant frequency shift due to the opened reflectors,

$$\frac{\Delta f}{f_0} = \frac{\phi_2 \Delta v / v_0 + \bar{\phi}_2 \bar{\Delta v} / v_0}{2\pi n - \phi_1 - \bar{\phi}_1} \quad (3)$$

Reflection From Two Cascaded Arrays

The new reflection coefficient of two cascaded arrays is calculated from the reflection and transmission coefficients of the shorted "B" array and opened "A" array shown in Figure (3).

REFLECTION COEFFICIENT OF TWO CASCADED ARRAYS

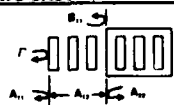


FIGURE 3

The new cascaded reflection coefficient is

$$\Gamma = A_{11} + \frac{A_{12}^2 B_{11}}{1 - A_{22} B_{11}} = \Gamma_0 e^{j\phi} \quad (4)$$

where

$$A_{11} = \pm |A_{11}| e^{j\alpha_{11}}$$

$$A_{12} = \pm |A_{12}| e^{j\alpha_{12}}$$

$$B_{11} = \pm |B_{11}| e^{j\beta_{11}}$$

From array symmetry and the particular choice of reference lines in Figure (3),

$$|A_{11}| = |A_{22}| \quad \text{and} \quad (5a)$$

$$\alpha_{22} = \alpha_{11} - \pi \left(1 + \frac{\Delta f}{f_0}\right) \quad (5b)$$

Equation (5) can be used to eliminate A_{22} from Equation (4).

A_{12} can also be removed from Equation (4) by using the following phase and amplitude relations resulting from conservation of energy in lossless reciprocal two-port junctions³

$$\alpha_{12} = \frac{\alpha_{11} + \alpha_{22} \pm \pi}{2} \quad (6)$$

$$|A_{11}|^2 + |A_{12}|^2 = 1 \quad (7)$$

Substituting Equation (5b) into Equation (6) gives the phase of the transmission coefficient in terms of the phase of the reflection coefficient

$$\alpha_{12} = \alpha_{11} - \frac{\pi \Delta f}{2f_0} \quad (8)$$

With these relations the expression for the cascaded reflection coefficient, Equation (4), becomes

$$\Gamma = \frac{A_{11} e^{j\alpha_{11}} [1 + (B_{11}/A_{11}) e^{j(\alpha_{11} + \beta_{11} - \pi \Delta f / f_0)}]}{1 + A_{11} B_{11} e^{j(\alpha_{11} + \beta_{11} - \pi \Delta f / f_0)}} \quad (9)$$

where the absolute value signs have been dropped for convenience.

Close to resonant frequency where small angle approximations of $\sin x \approx x$ and $\cos x \approx 1$ are valid, the magnitude and phase of Equation (9) become

$$\begin{aligned}\phi &= \alpha_{11} + \left[\frac{B_{11}}{A_{11} + B_{11}} - \frac{A_{11} B_{11}}{1 + A_{11} B_{11}} \right] \\ &\quad \cdot (\alpha_{11} + \beta_{11} - \frac{\pi \Delta f}{f_0})\end{aligned} \quad (10)$$

$$\Gamma_0 = \frac{A_{11} + B_{11}}{1 + A_{11} B_{11}} \quad (11)$$

The next step is to derive the magnitude and phase of the single A and B arrays to use in these cascaded array equations.

Shorted Array B

Approximations for the reflection phase and amplitude of a single array can be derived from reference 2 as was done in reference 4. If the impedance mismatch ratio in shorted array B is $z_b = z_{b2} / z_{b1}$ where z_{b2} and z_{b1}

are impedances of the line and gap, respectively, n_b = number of reflector sections, and the line and gap transmission length are equal, the amplitude of the reflection coefficient is²

$$B_{11} = \frac{z_b^{2n_b} - 1}{z_b^{2n_b} + 1} \quad (12)$$

and the phase of the reflection coefficient can be shown to be

$$\beta_{11} = \left[\frac{\pi}{2} - \frac{\pi}{2(z_b^2 + 1)} - \frac{4z_b}{(z_b + 1)^2} S(n_b, z_b) \right] \frac{\Delta f}{f_0} \quad (13)$$

where

$$S(n_b, z_b) = \frac{\sinh[(n_b - 1)z_b']}{\sinh(n_b z_b')}$$

and

$$z_b' = \frac{z_b - 1}{\sqrt{z_b}}$$

For small impedance mismatches ($z_b \approx 1$), Equation (13) reduces to

$$\beta_{11} = \left[\frac{\pi}{2} - \frac{\pi}{1 - S(n_b, z_b)} \right] \cdot \frac{\Delta f}{f_0} \quad (14)$$

The $\pi/2$ in Equation (14) arises from the choice of reference plane at the edge of the reflector line. Equations (13) and (14) are similar to equations in reference 4 except here the sign of the reflection coefficient is explicit in the amplitude term instead of the phase term, and small Δf approximations are assumed.

Opened Array A

The reflection coefficient amplitude of the opened array A is similar to Equation (12)

$$A_{11} = \frac{z_a^{2n_a} - 1}{z_a^{2n_a} + 1} \quad (15)$$

where $z_a = z_{a2}/z_{a1}$ and n_a is the number of opened strips. Note that z_a may not equal z_b if the impedance ratios are different for opened and shorted reflectors as they are for the quartz resonators discussed later.

The expression for the phase of opened array A is similar to Equation (14) except that the center frequency of array A does not occur at f_0 , but due to the velocity shift through the opened array ($\Delta v = v - v_0$), the center frequency is given by $f_0' = f_0(1 + \Delta v/v_0)$. Thus, the phase of the opened array A for small $\Delta v/v_0$ becomes

$$\alpha_{11} = \left[\frac{\pi}{2} - \frac{\pi}{1 - S(n_a, z_a)} \right] \cdot \left(\frac{\Delta f}{f_0} - \frac{\Delta v}{v_0} \right) \quad (16)$$

Combined Arrays

Substitution of Equations (12), (14), (15) and (16) into Equations (10) and (11) gives the expressions for phase and amplitude of the cascaded opened and shorted arrays

$$\phi_1 = \left[\frac{\pi}{2} - \frac{\pi}{1 - S(n_a, z_a)} \right] \cdot \left(1 + \frac{B_{11}}{A_{11} + B_{11}} - \frac{A_{11}B_{11}}{1 + A_{11}B_{11}} \right) + \left[-\frac{\pi}{2} - \frac{\pi}{1 - S(n_b, z_b)} \right] \cdot \left(\frac{B_{11}}{A_{11} + B_{11}} - \frac{A_{11}B_{11}}{1 + A_{11}B_{11}} \right) \quad (17)$$

$$\phi_2 = -\left[\frac{\pi}{2} - \frac{\pi}{1 - S(n_a, z_a)} \right] \cdot \left(1 + \frac{B_{11}}{A_{11} + B_{11}} - \frac{A_{11}B_{11}}{1 + A_{11}B_{11}} \right)$$

$$\Gamma_0 = \frac{z_a^{2n_a} z_b^{2n_b} - 1}{z_a^{2n_a} z_b^{2n_b} + 1} \quad (18)$$

The factors in Equation (17) involving the amplitude terms can be simplified if $z \approx z_b$ and the total reflection amplitude $|\Gamma_0| \approx 1$, as is true for the resonators discussed here. With these approximations, Equation (17) becomes

$$\phi_1 = \frac{\pi}{2} - \frac{\pi}{1 - S(N, z_b)} \quad (19)$$

and

$$\phi_2 = -\left[\frac{\pi}{2} - \frac{\pi}{1 - S(n_a, z_a)} \right] \cdot \frac{1 + |B_{11}|}{1 + A_{11}B_{11}}$$

where $N = n_a + n_b$. Note that ϕ_1 , the phase term accounting for propagation distance to the effective center-of-reflection within the array and back, is now a constant independent of the number of opened and shorted reflector strips. This is not valid if the impedance discontinuity changes significantly as the reflectors are opened since, in that case, the effective center-of-reflection changes.

The frequency shift of a resonator having one cascaded reflector array characterized by Equation (19) and the other reflector array characterized by similar expressions is then given by Equation (3), repeated here,

$$\frac{\Delta f}{f_0} = \frac{\phi_2 \Delta v/v_0 + \bar{\phi}_2 \bar{\Delta v}/v_0}{2\pi n - \phi_1 - \bar{\phi}_1} \quad (20)$$

Equation (20) is compared with experimental data in the next section of this paper.

Experimental Data

Series Resonance Shift

The 155.782 MHz SAW resonators used in this work were fabricated of aluminum on ST quartz. The resonators had a cavity size of 41.5 wavelengths which contained a 40 split-finger pair transducer. The aluminum thickness of the reflectors was 3000Å. The resonator was placed in a balanced hybrid circuit and the minimum loss resonant frequency was monitored on a spectrum analyzer together with a frequency counter as the reflector arrays were opened using a xenon laser.

Figure 4 shows a plot of the shift in series resonance frequency in parts per million against the number of reflectors opened in the first array. Here z_a and z_b are the line to gap acoustic impedance ratios for the opened and shorted reflector sections. The change in velocity from the opened to shorted sections $\Delta v/v_0$ was extracted from the work of Marefield and Tournois⁶. The theoretical curve followed experimental data to within measurement tolerance. The shift due to entirely opening the first reflector array was 90 ppm.

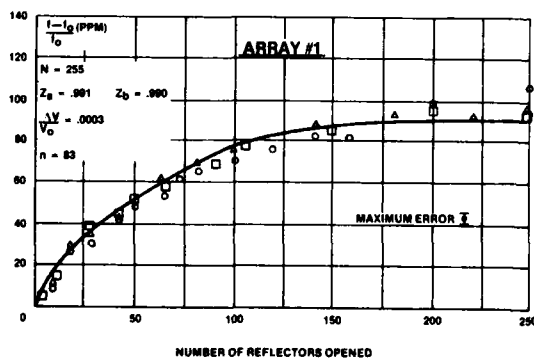


FIGURE 4

Figure 5 is a plot of resonance shift achieved by opening the fingers of the second array where array one is already entirely open. The average observed shift due to opening both gratings was 180 ppm. In this plot the experimental data fit reasonably well for two out of three resonators shown but there was a significant deviation for the third device. In this resonator the transverse modes began to interfere with the main mode - this effect has been correlated with spectrum analyzer photographs taken during tuning and will be discussed in more detail later.

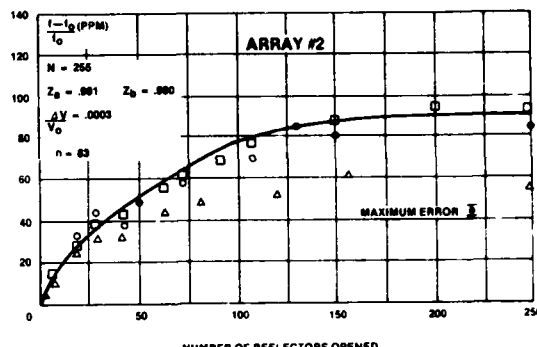


FIGURE 5

Both the theoretical and experimental work show that the series resonance frequency changed most rapidly when the reflectors close to the cavity are opened. Figure 6 plots the resonance frequency shift against number of fingers open per side; in this case there was symmetry about the cavity center as to the number of fingers opened and shorted. Here the data again fit well to the theoretical until there were some 150 reflectors opened per side, at which point the data again began to pull away from the theoretical curve because of transverse mode interference.

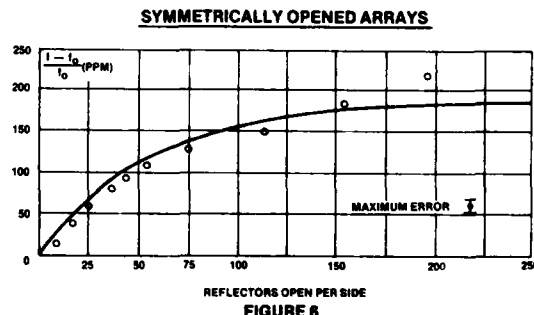


FIGURE 6

Transverse Modes

Figure 11 is a series of spectrum analyzer photographs illustrating at progressive intervals the interference of transverse modes with the main resonance mode during the operation of laser scribing. The transducer in the cavity of this resonator was a 40 wavelength uniform overlap split-finger type, which, because of its strong coupling to the transverse mode, made it a good candidate for observing these modes. As tuning progressed, the main mode moved up towards the transverse modes. The transverse modes remained relatively fixed, but began to shift slightly when the main mode closely approached them.

At this time we also began to see the deviation in resonance frequency from the theoretical curve as was noted earlier.

It appeared in part 4 of Figure 11 that the first transverse mode dominated resonance after some 300 fingers were opened which was in agreement with what Staples⁷ observed in 1975. The last insert of Figure 7 shows the heightened transverse mode activity characteristic of entirely open aluminum reflector arrays.

Two Pole Resonator

A two pole resonator whose center reflector had 150 elements and outer reflectors had 300 elements was fabricated with shorted reflector arrays throughout. The configuration of this aluminum on ST quartz, 155 MHz, two pole resonator is shown at the top of Figure 12. Also shown here are spectrum analyzer photographs of this resonator in the 50 ohm system before impedance matching. This figure shows the resonator at progressive intervals of laser tuning. The average shift in resonant frequency, as defined at top of this figure, was noted after each interval. Photograph 1 (entirely shorted reflector arrays) shows the characteristic two peaked response for the over-coupled two-pole resonator and also exhibits a number of transverse modes. Both cavity transducers were unapodized. After opening all reflector arrays a shift of 208 ppm was observed which was approximately 20% higher than what was predicted by a full two pole resonator model including transducer effects.

This part of the experimental work also proved out another advantage to this type of tuning. The two pole resonator here was mounted in a completely sealed package having a glass lid. A xenon laser was then focused through this lid to open the metal reflector elements, thereby correcting for frequency shifts caused by package sealing.

Resonator Series Resistance

Since the impedance mismatch of the opened reflectors is less than that of the shorted reflectors, the magnitude of the reflection coefficient will decrease as the array is opened. Consequently, the resonator characteristics, as described by its equivalent circuit (reference 5), will vary. For example, the decrease in Γ , given by Equation (18), increases series resistance R_1 . The series resistance for a resonator with identical reflectors⁵ is

$$R_1 = R_a \frac{1 - |\Gamma|}{1 + |\Gamma|} \quad (21)$$

where R_a is the acoustic radiation resistance of the isolated transducer. Following the same procedure described in reference 5, it can be shown that if the reflectors are not identical such that $\Gamma \neq \bar{\Gamma}$, the series re-

sistance is

$$R_1 = R_a \frac{1 - |\Gamma| |\bar{\Gamma}|}{(1 + |\Gamma|)(1 + |\bar{\Gamma}|)} \quad (22)$$

with

$$\Gamma = \frac{z_a^{2n_a} z_b^{2n_b} - 1}{z_a^{2n_a} z_b^{2n_b} + 1}$$

The change in series resistance as measured by changes in insertion loss were plotted for both the theoretical case and for experimental units in Figure 7. The vertical axis gives series resistance in ohms and the horizontal axis again indicates the number of strips opened by xenon laser.

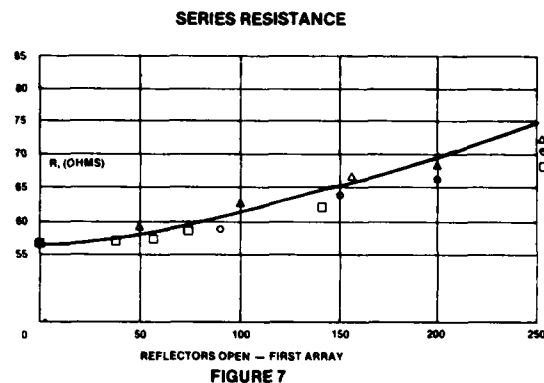


FIGURE 7

The experimental data fit was quite good until the last part of the array was opened. At this point the attenuation of the wave through the reflectors diminished the change in reflection coefficient resulting from opening a reflector short in this part of the array. Note that the slope of the theoretical resistance curve in Figure 7 increased slightly as the reflector array was opened from the cavity side, while the slope of the resonant frequency shift decreased as can be seen in Figures 4-6. The resonant frequency depends on the phase of the reflection coefficient for the combined opened and shorted arrays. Since the rate of the phase shift diminishes as reflectors are opened further from the cavity end of the array, the rate of the frequency shift also diminishes and approaches the asymptotic limit of a completely open array. However, as shown in Equation (22), the series resistance depends only on the magnitude of the reflection coefficient, not on the phase. Since the opened reflectors have a slightly smaller impedance discontinuity than the shorted reflectors, the reflection coefficient decreases and consequently R_1 increases as the array is opened. But opening a reflector on the cavity side of the array has approximately the same effect

on the magnitude of the reflection coefficient as opening a reflector on the far side of the array, so the series resistance curve is approximately linear. The fact that there is a slight increase in slope is due to the reduced total reflection coefficient of the opened reflectors which increases the effect of opening the remaining shorted reflectors.

Oscillator Measurements

All previous measurements were made using a spectrum analyzer together with frequency counter. This set up was useful because it gave not only series resonant frequency as measured by minimum loss point, but also showed the effects of transverse modes on main mode during tuning. It also allowed for measurement of R_1 by the change in insertion loss. This section gives data for the shift in oscillator frequency of an oscillator network controlled by a SAW undergoing laser tuning.

Figure 8 shows the oscillator circuit used here. Figure 9 is a plot of oscillator frequency shift as noted on a frequency counter against number of fingers opened in the first reflector array. Here we found the data from the experimental units fell significantly above the theoretical curve.

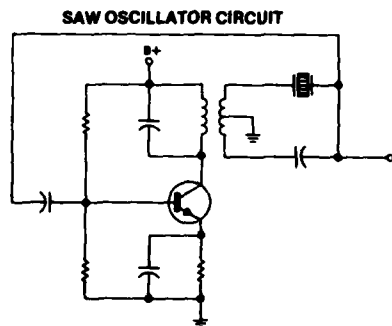


FIGURE 8

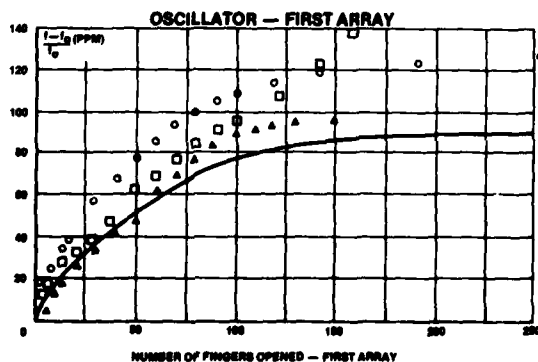


FIGURE 9

The reason for that was that the theoretical curve gives the true series resonance as measured by minimum loss point, whereas oscillator frequency is not only determined by shift in resonant frequency of the resonator, but also by changes in the phase slope of the resonator due to reduced reflection coefficient which is not included in the theoretical curve.

Fabrication Tolerances

A batch of fifty resonators fabricated from ten different wafers were measured in the course of these laser tuning experiments. The standard deviation of the resonant frequencies of these devices is tabulated in Figure 10. The average standard deviation for the devices row to row was significantly smaller than the average standard deviation of wafer to wafer changes. This fact alone indicated some substantial contribution to fabrication tolerances resulting from the photomask fabrication effects. But this data does point out that it will be possible to adjust for approximately 50% of the total fabrication related shifts in resonance frequency by laser burning of shorted metal reflector arrays.

STANDARD DEVIATION CENTER FREQUENCY

WITHIN ROWS	± 41 PPM
WITHIN WAFERS	± 142 PPM
WAFER TO WAFER	± 167 PPM

FIGURE 10

Conclusions

An accurate useful model has been developed for the frequency shift and series resistance change as the elements of a shorted metal reflector array are opened. Experimental data has been presented demonstrating an overall resonance shift of 180 ppm for completely open to completely shorted array. The series resistance was shown to increase by 12% for single pole resonators having one array entirely open. It has been demonstrated that it is practical to use a laser to accomplish this tuning on active one pole and two pole resonators and resonators in oscillator circuits. It has also been shown that resonators can be tuned In Situ by focusing a laser through a glass covered package, thereby also correcting for frequency shifts caused by sealing. Furthermore, the affects of tuning on transverse mode spectra were observed which suggests that

tuning method could be used in the study of transverse modes in SAW resonators.

Acknowledgements

The authors wish to acknowledge the contributions made by the following individuals: R. Kinsman for technical discussions, P. Fenlon for help in testing and evaluation, J. Spangler for the oscillator circuit design, J. Guertin, R. Ackerman and A. Klebba for graphics, and J. Rieckenberg for the preparation of the manuscript.

References

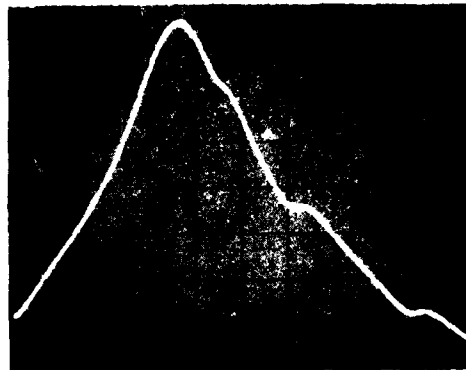
- 1) F. G. Marshall, "Surface Acoustic Wave Resonators Constructed of Aluminum on ST Quartz for Use in High Stability Feedback Oscillators", 1975 Ultrasonics Symposium Proceedings, pp. 290-292 Cat. #75 CHO 994-4SU
- 2) E. K. Sittig and G. A. Coquin, "Filters and Dispersive Delay Lines Using Repetitively Mismatched Ultrasonic Transmission Lines", IEEE Trans. Sonics and Ultrasonics, SU-15, No. 2, pp. 111-119, April 1968
- 3) See for example, H. J. Carlin, "The Scattering Matrix in Network Theory", IRE Trans. Circuit Theory, pp. 88-97, June 1956
- 4) P. S. Cross, "Properties of Reflective Arrays for Surface Acoustic Resonators", IEEE Trans. Sonics, SU-23, No. 4, pp. 255-262, July 1976
- 5) J. S. Schoenwald, R. C. Rosenfeld, and E. J. Staples, "Surface Wave Cavity and Resonator Characteristics - VHF to L-Band", 1974 Ultrasonics Symposium Proceedings, pp. 253-256, Cat. #74 CHO 896-1SU
- 6) C. Maerfeld, P. Tournois, Electronics Lett., Vol. 9, p. 115, 1973, as referenced in W. H. Haydl, B. Dischler, and P. Hiesinger, "Multimode SAW Resonators - A Method to Study the Optimum Resonator Design", 1976 Ultrasonics Symposium Proceedings, pp. 287-296, Cat. #76 CH 1120-5SU
- 7) E. J. Staples and R. C. Smythe, "Surface Acoustic Wave Resonators on ST-Quartz", 1975 Ultrasonics Symposium Proceedings, pp. 307-310, IEEE Cat. #75 CHO 994-4SU

TRANSVERSE MODES — TUNING ONE ARRAY

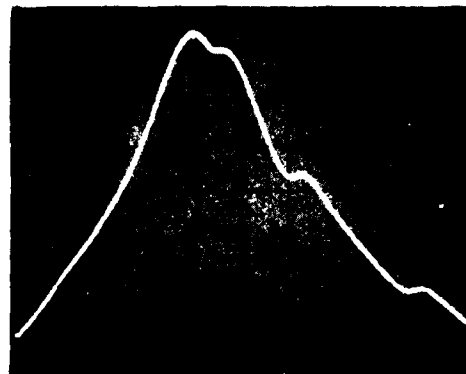
VERTICAL 2 db/DIV

HORIZONTAL 20 kHz/DIV

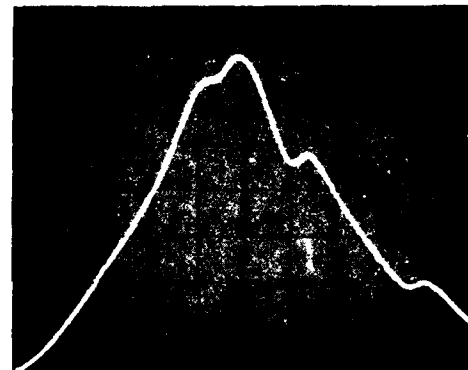
LOG REF. -6 db



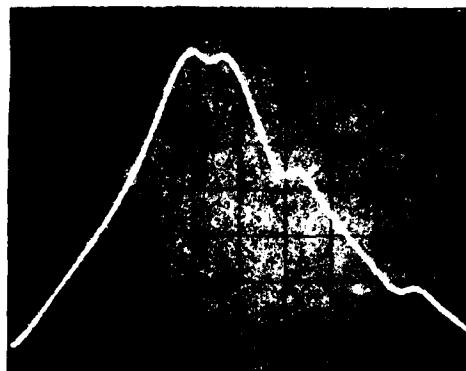
1 — BEFORE TUNING
 $f_0 = 154.620$



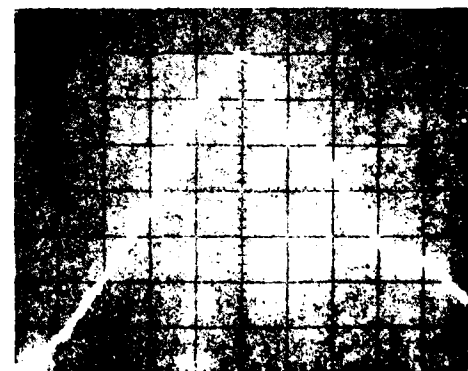
2 — 80 REFL. OPEN RIGHT ARRAY



4 — ARRAY 1 OPEN
30 REFL. OPEN ARRAY 2



3 — ARRAY 1 OPEN



5 — ALL REFL. OPEN
 $f_0 = 154.634$

FIGURE 11

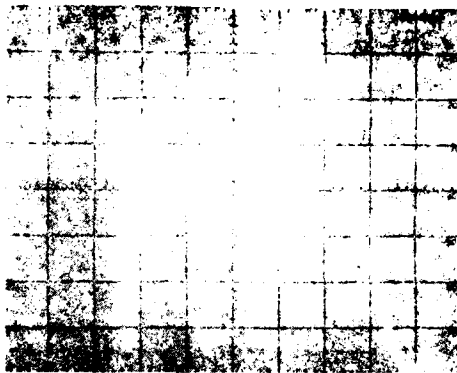
2 POLE RESONATOR

2 db/DIV; 20 kHz/DIV

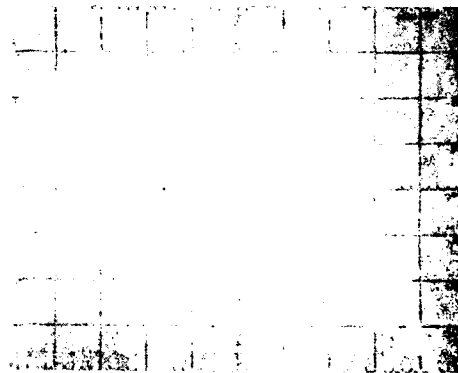
REF -5 db (1, 2, 3)
-6 db (4)

$$\frac{\Delta f_{AV}}{f_{AVO}} = \frac{f_{AV} - f_{AVO}}{f_{AVO}} \times 10^4$$

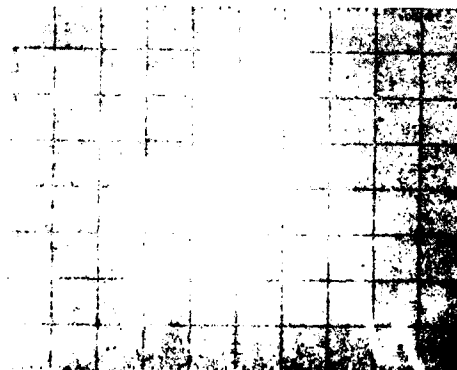
$$f_{AV} = \frac{f_1 + f_2}{2}$$



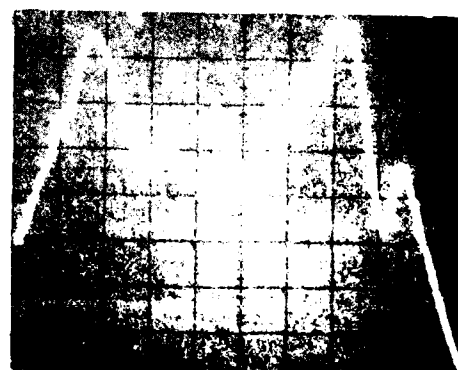
1 — BEFORE TUNING
 $\Delta f/f:0$



3 — RIGHT, LEFT ARRAYS OPEN
 $\Delta f/f:123$



2 — RIGHT ARRAY OPEN
 $\Delta f/f:65$



4 — ALL REFLECTORS OPEN
 $\Delta f/f:208$

FIGURE 12

TUNABLE VARIABLE BANDWIDTH/FREQUENCY SAW RESONATORS *

Jeff Stuart Schoenwald
Teledyne MEC
Palo Alto, California 94304

Abstract

A SAW resonator device is demonstrated such that 3 dB fractional bandwidths of .1% to .05% can be achieved in the same device prior to implementation of input/output matching networks. Such matching networks have a deleterious effect on shape factor and out-of-band rejection. The technique used is one in which two or more resonant SAW cavities are acoustically coupled in-line with reflectors between cavities controlling the degree of coupling. If the coupling reflectors are formed as large interdigital transducers, then the degree of coupling and consequent degeneracy splitting may be controlled by loading the coupling reflector/transducer with a suitable impedance. Input/output transducers can then be matched for the bandwidth and characteristic impedance of the coupled system, reducing mid-band distortion and optimizing spurious suppression and low shape factor. Forming end reflectors as extended transducers and properly loading these allows for additional degrees of freedom in tuning bandwidth and center frequency. Electrically cascading two or more such "monolithic" SAW crystal filters provides additional spurious rejection. Thus, it is shown that a variable bandwidth/frequency tunable monolithic SAW resonator is feasible.

Summary

A SAW resonator, fabricated on ST-quartz, using metalization loading, is shown to possess variable bandwidth and center frequency capability at the post-fabrication stage of manufacturing. By designing the reflector as an interdigital transducer (IDT) with many finger pairs, it may be loaded by a reactive or resistive element.¹ The electrical phase and magnitude of the resulting composite has a highly variable effect on the phase and magnitude of reflected surface acoustic waves incident on the reflector structure. In a resonant cavity structure, this results in a dramatic change in the frequencies of cavity modes--enabling one to control the separation of resonance frequencies and, hence, the filter bandwidths. For the same reasons, but to a somewhat lesser extent, the center frequency can be tuned. The use of such variable reflectors in multi-pole, SAW resonator filters results, effectively, in a SAW equivalent of the monolithic crystal filter with post-fabrication frequency and bandwidth external tunability--the tunable monolithic SAW crystal filter. Two-pole and three-pole filter examples will be demonstrated.

The Tunable Reflector

Figure 1a is a schematic representation of transmission between two wideband, SAW probe IDT's through a reflector fabricated as a large IDT with many finger pairs to which has been connected a shunt inductor. We may specify a simple equivalent circuit. For this reactively-loaded transducer (Figure 2), and write down its electrical impedance Z (neglecting parasitic resistance), as measured across the IDT:

$$Z = \frac{(\omega L)^2 R + j\omega L R (1 - \omega^2 LC)}{R^2 (1 - \omega^2 LC)^2 + (\omega L)^2} \quad (1)$$

where R = acoustic radiation resistance of large reflecting IDT, C = static capacitance, and L = variable shunt inductance.

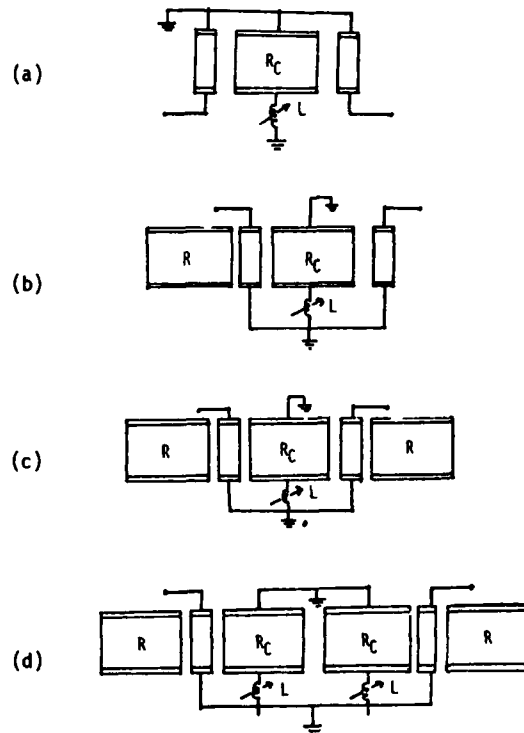


Figure 1. Various device configurations discussed in text.

- (a) Transmission through a tunable reflector R_C .
- (b) Transmission between an external IDT and one located in a cavity formed by two reflectors - R_C , which is tunable, and R , which may also be tuned, if desired.
- (c) Two-pole, two-port resonator.
- (d) Three-pole, two-port resonator.

The real and imaginary parts of Z are shown in Figure 3a, and the magnitude of Z in Figure 3b as a function of L for typical values of R and C , with L varying over a range such that L and C are parallel resonant in the vicinity of the cavity resonant frequency.

It should be readily deduced that the phase of reflected waves due to regeneration vary as L is varied. This is to be distinguished from waves reflected due to mass loading, which is not affected by the tuning process.

* This research was supported by the Naval Air Development Center, Warminster, Pa. 18974, under Contract No. N62269-76-C-0417; Elliott L. Ressler, Program Engineer.

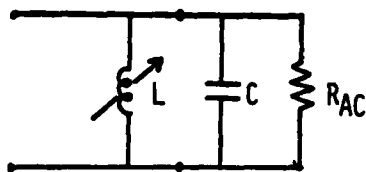


Figure 2. Simple equivalent circuit representation of a SAW IDT loaded by a parallel inductor as viewed from the electrical port. R_{AC} = acoustic radiation resistance at synchronism, C = IDT static capacitance, L = variable parallel inductance.

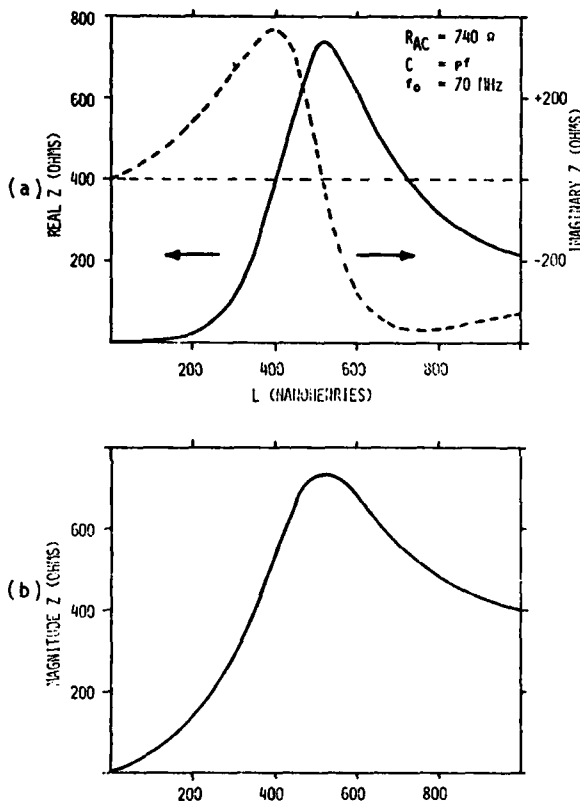


Figure 3. (a) Real and imaginary impedance, and (b) impedance magnitude versus parallel inductance of schematically represented transducer shown in Figure 2.

Furthermore, because resistive dissipation will also reach a maximum at peak reflection, it does not necessarily follow that reflection is enhanced, although transmission attenuation does change dramatically, as shown in Figure 4.

If reflection were enhanced as a result of resonantly tuning the reflector, then the resonance resistance, r , as shown in the two-port (single-pole) equivalent circuit² in Figure 5, would decrease as:

$$r = R_0 \frac{1 - |\Gamma|}{1 + |\Gamma|} \quad (2)$$

where $|\Gamma|$ = fractional amplitude magnitude of the surface acoustic wave reflected from one of the reflectors forming a resonant cavity, and R_0 = tap IDT

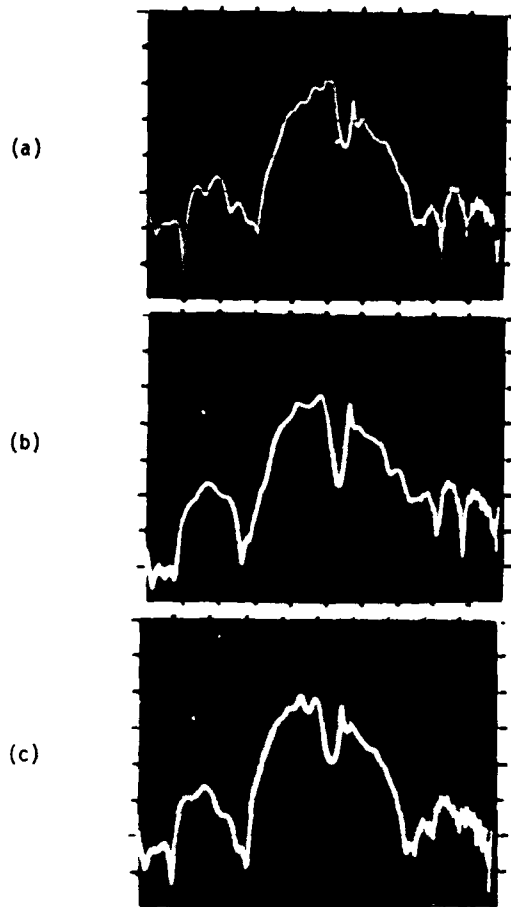


Figure 4. Transmission through a tuned reflector R_C as parallel inductance increases (see Figure 1a). (a) Minimum inductance. (b) Inductance tuned for LC resonance at band center. (c) Maximum inductance.

Vert: 10 dB/Div. Horiz: 1 MHz/Div.

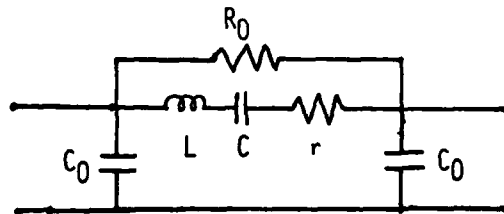


Figure 5. Two-port, single-pole lumped element equivalent circuit. R_0 = input/output IDT acoustic radiation resistance at synchronism; C_0 = input/output IDT static capacitance; L , C = motional inductance and capacitance; r = parasitic resistance due to electrical and material dissipation losses.

acoustic radiation resistance.

If this were the case, then resonance insertion loss would decrease as reflector efficiency increases by virtue of reactive tuning. Experimentally, this is not the case, as will be shown.

Tuning the Single-Pole Resonator

Consider the structure in Figure 1b. As the reflector, R_c , is inductively tuned through resonance at band center, a resonant mode appears whose frequency decreases as the tuning inductance increases (Figure 6). The point at which maximum transmission attenuation occurs in R_c (Figure 4b) corresponds to a condition midway between those observed in Figures 6b-c. Note that overall resonance signal strength under these circumstances has decreased, with consequent reduction in rejection from the acoustic transversal signal level. The resonance frequencies in Figures 6a and 6d are identical, although mode coupling occurs in 6b and 6c, splitting the degeneracy as the two interact strongly. It is concluded that the mode whose frequency remains unchanged at the two extreme values of inductance L is probably the mode defined by impedance conditions due to mass loading, while the resonance that shifts with changing L is the regenerative mode controlled by tuning the electrical impedance of the reflector, R_c . The earlier statement that resonance signal strength is not enhanced by reactively tuning the reflector is thus demonstrated.

Tuning the Two-Pole Resonator

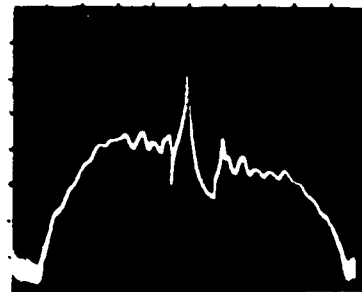
Next, consider the device in Figure 1c. The reflector, R_c , creates, effectively, two cavities that are coupled through R_c in much the same way that bound quantum states in two weakly coupled, finite square well potentials interact to form split degeneracy state pairs. This device has, in fact, been presented earlier³, but its tunable bandwidth capability was not then exploited. By varying the inductive load over its full range, we may obtain the results shown in Figure 7. Observe that the lower frequency resonance suffers no shift, while the upper one appears to breath accordion fashion as L increases. The sawtooth behavior appearing on the right side of the resonance peaks is due to transverse modes previously discussed and occurs when the input/output transducers are not properly apodized to suppress all but the fundamental mode transverse to the direction of propagation. Figure 8 shows the device behavior at the two extreme values of L .

Figure 9 shows the device behavior when the reflector, R_c , and inductor, L , are resonant at the band center. The mode in the center does not shift with L and has been identified with mass loading boundary conditions. Two modes, approximately .25 MHz apart, form "shoulders" at peak resonance, and move down in frequency as L increases, but maintain a fixed separation. By partially matching the impedance of the input/output IDT's with simple series inductances, it is possible to produce a flat response (apart from transverse modes) across the bandwidth defined by the separation of the two modes. Figure 10 shows such a matched and tuned device with 3 dB insertion loss.

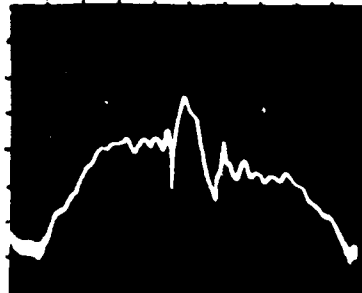
The Three-Pole Resonator

We may now consider the next level of complexity: the three-pole monolithic bandwidth tunable SAW crystal filter. Figure 1d is representative of some of the increased numbers of degrees of freedom of L . The outer reflectors are shown without tuning inductors, but may be dynamically loaded as well.

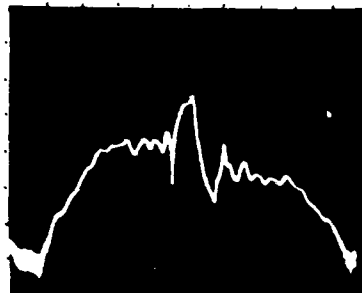
(a)



(b)



(c)



(d)

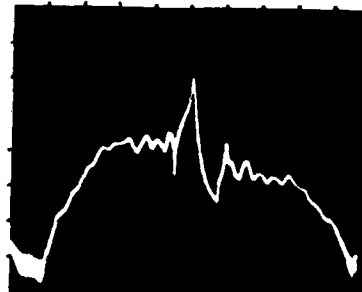


Figure 6. Two-pole resonance behavior as R_c is inductively tuned. L increases with successive photographs. Vert: 10 dB/Div. Horiz: .5 MHz/Div.

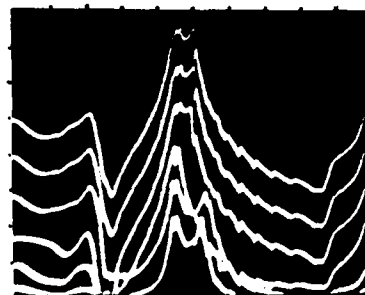


Figure 7. Mode variation with increasing L . Successive traces shifted down by 10 dB. Vert: 10 dB/Div. Horiz: .1 MHz/Div.

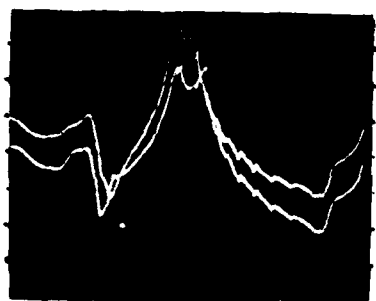


Figure 8. Top: Mode structure for minimum L value. Bottom: Mode structure for maximum L value. Vert: 10 dB/Div. Horiz: .1 MHz/Div.

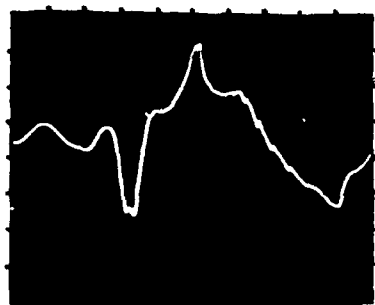
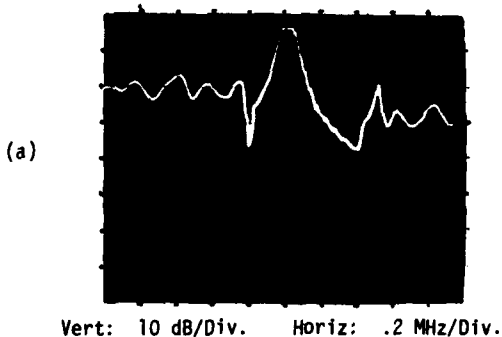


Figure 9. Frequency response of two-pole resonator with static capacitance of reflector R_c resonantly tuned by parallel inductor at center frequency. Vert: 10 dB/Div. Horiz: .1 MHz/Div.



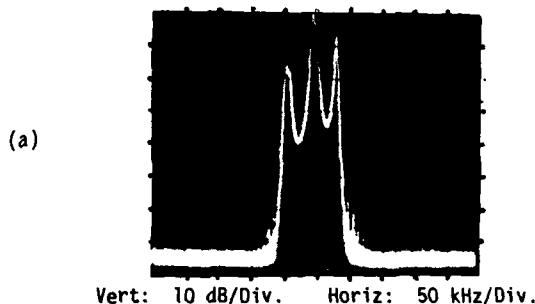
Vert: 10 dB/Div. Horiz: .2 MHz/Div.



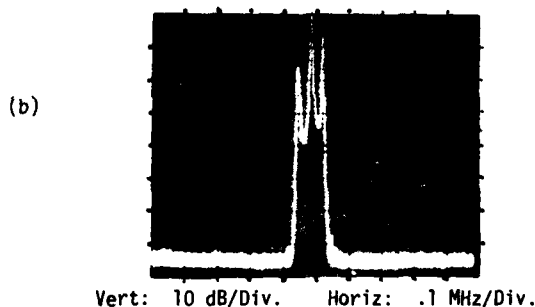
Vert: 2 dB/Div. Horiz: 50 kHz/Div.

Figure 10. Frequency response of tuned, matched, two-pole, two-port resonator.

In the absence of any tuning elements, either across the reflectors or the input/output IDT's, we obtain the response in Figure 11. In this device, the tap IDT's have been cosine weighted to suppress transverse mode structure, which has been shown to mar the smoothness of the passband response (see Figures 7 and 10b). Tuning one or both of the two coupling reflectors has much the same effect as was achieved in the two-pole device. Tuning two or more reflectors in the array has the effect mainly of controlling the interaction of mode coupling to provide a passband with uniform and minimum ripple. Resonating the tap transducers with a series inductor further reduces insertion loss and passband ripple. But, at the same time, matching reduces rejection of the transducer passband response from the resonant signal. Figure 12 illustrates the bandwidth variation that can be achieved in a device that is both matched and reflector tuned. Here, the passband response has been tuned from 70 kHz to 40 kHz bandwidth (3 dB) by a smooth variation in the inductances loading the coupler reflectors, R_c . More dynamic behavior is achieved by tuning end reflectors as well. In this fashion, bandwidths between 18 and 85 kHz have been achieved.



Vert: 10 dB/Div. Horiz: 50 kHz/Div.



Vert: 10 dB/Div. Horiz: .1 MHz/Div.

Figure 11. Frequency response of untuned, unmatched, three-stage cascaded, three-pole resonator. Insertion loss = 26 dB.

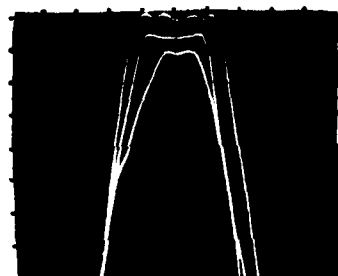


Figure 12. Representative example of bandwidth control achievable in the same device by tuning reflector impedance. Vert: 2 dB/Div. Horiz: 20 kHz/Div.

Perhaps the most effective way to eliminate, or sharply reject, nonresonant acoustic feedthrough is to devise transverse acoustically coupled resonators such as first described by Tiersteg and Smythe,⁴ and illustrated by Staples and Smythe.⁵ The design format of the two- and three-pole SAW filters shown here compel us to examine the cascaded crystal arrangement to improve rejection of undesirable acoustic coupling away from the passband response. This leads us to a discussion of the cascaded in-line device configuration and its performance as an N-pole Butterworth filter.

Cascaded SAW Resonator Filters

As Li, et al.⁶ have shown theoretically by application of a transmission line analysis of various in-line, multi-cavity designs, the shape factor decreases in relation to the increased number of cavities, or poles, included in the configuration. An N-pole Butterworth filter doubles in bandwidth with every 6N dB decrease from peak signal strength. As a demonstration of this behavior, we offer the results of a 70 MHz filter constructed from a cascaded triplet of three-pole resonators. Input and output transducer taps have simple series L inductances, thus impedance matching is not optimized. The electrical coupling between crystals is via series or shunt inductors with no discernable difference being evident. Figure 13 is a series of spectrum analyzer oscilloscope traces showing the broad and narrow band frequency characteristics of this device. The ± 1 dB ripple in the passband could have been reduced both through more extensive reflector tuning (not all of which were loaded) and better impedance matching at the input/output ports. The derivative of a transition slope is quite steep, exceeding 2000 dB/MHz on the high frequency transition skirt. Because of the additional pole zero behavior of the interstage and matching network components, the device satisfies an effective 11-pole Butterworth filter requirement. Average insertion loss is 13 dB at midband due to incomplete matching and parasitic loss encountered in the metal electrodes and inductors. Close-in dynamic range exceeds 90 dB, but is ultimately limited to approximately 42 dB from in-line acoustic feedthrough. Beyond the passband response of the tap transducers, spurious acoustic response is below -75 dB from dc to 1 GHz.

Center Frequency Tuning

Because the mass loading boundary conditions are not altered by any electrical boundary conditions that may vary as a function of electrical loading, the lower end of the resonance passband is not as tunable as might at first be expected. Nevertheless, given the number of elements in a multi-pole device that can be tuned, it has been observed that in the devices shown here that the center frequency can be tuned by almost half the maximum achievable 3 dB bandwidth of the device. Thus, a frequency shift of 30 to 35 kHz has been seen in filters possessing a maximum bandwidth of 70 kHz.

Conclusion

We have demonstrated the effect of reactive loading on an extended IDT functioning as a SAW reflector and its application to SAW resonator filters. In particular, the phase of the reflected wave may be externally controlled to vary the bandwidth of a multi-cavity, coupled mode filter. Examples of two- and three-pole devices are presented and a cascaded three-stage, three-pole filter is shown which, when matched at both ports, demonstrates 11-pole Butterworth filter characteristics. The external control of reflective behavior of the separate arrays of a device can affect the center frequency as well.

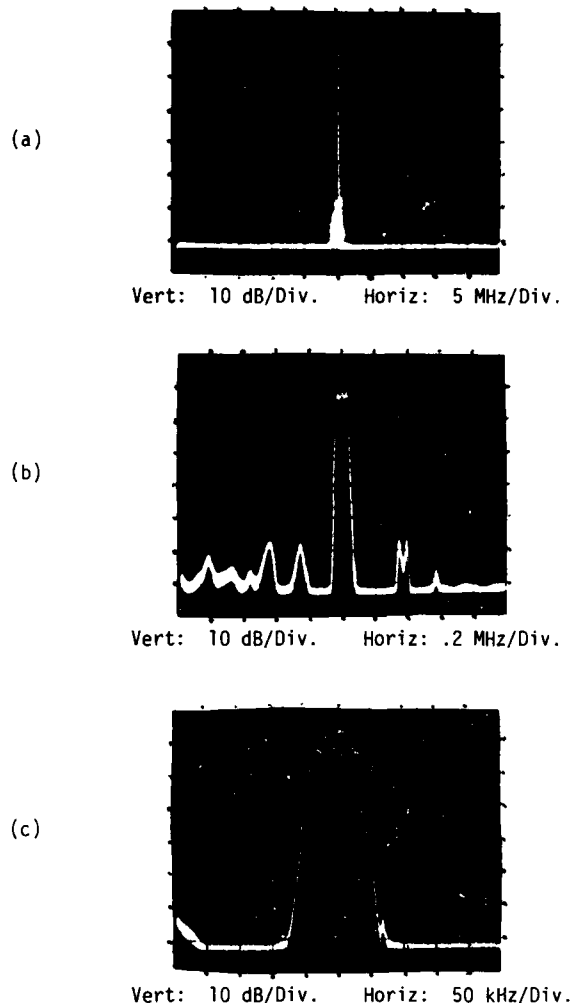


Figure 13. Tuned matched response of three-stage, three-pole SAW resonator filter. Insertion loss = 13 dB.

Acknowledgements

The author wishes to acknowledge the contributions made by the following individuals: C. Grasse and K. Braun for useful discussions and technical assistance, D. Cooper for testing and evaluation, T. Volcjak for device fabrication, J. Ho and P. Zoboli for graphics, and E. Steel for preparation of the manuscript.

References

1. H.M. Gerard, "Acoustic Scattering Parameters of the Electrically Loaded Interdigital Surface-Wave Transducer," IEEE Trans. on Microwave Theory and Techniques, MTT-17, pp. 1045-1046, November 1969.
2. W.R. Shreve, "Surface-Wave Two-Port Resonator Equivalent Circuit," 1975 IEEE Ultrasonics Symposium Proceedings, IEEE Pub. #75 CHO 994-4SU, 1975, pp. 295-298.

3. J. Schoenwald, "Optical Waveguide Model for SAW Resonators," in Proc. 30th Annual Frequency Control Symposium (U.S. Army Electronics Command, Ft. Monmouth, N.J.), pp. 340-345, June 1976.
4. H.F. Tiersten and R.C. Smythe, "Guided Surface-Wave Filter," Proc. 1975 IEEE Ultrasonics Symposium, IEEE Pub. #75 CHO 994-4SU, 1975, pp. 293-294.
5. E.J. Staples and R.C. Smythe, "SAW Resonators and Coupled Mode Filters," in Proc. 30th Annual Frequency Control Symposium (U.S. Army Electronics Command, Ft. Monmouth, N.J.), pp. 322-327, June 1976.
6. R.C.M. Li, J.A. Alusow and R.C. Williamson, "Surface-Wave Resonators Using Grooved Reflectors," in Proc. 29th Annual Frequency Control Symposium (U.S. Army Electronics Command, Ft. Monmouth, N.J.), pp. 167-176, May 1975.

DEEPLY ETCHED S.A.W. RESONATORS

Charles A. Adams
John A. Kusters
Hewlett-Packard Laboratories
Palo Alto, California

Summary

Recent development of surface acoustic wave devices as high frequency, high Q, one- and two-port acoustic resonators has created a new interest in such devices.¹⁻⁴ The surface wave resonator (SWR) consists of an interdigital transducer (IDT) contained within an acoustic resonant cavity. A pair of extended gratings consisting of deposited film lines,¹⁻² ion-implanted lines,³ or etched grooves³⁻⁴ form the cavity.

Early work on lithium niobate SWR's indicated significant conversion from surface waves to lossy bulk waves when grooved arrays were etched greater than 0.02 wavelengths deep.² Since then, every reported result on etched groove resonators has been for devices with shallow grooves less than 0.02 wavelengths deep.

This paper reports on work done on deeply etched (0.05 wavelengths or deeper) quartz SWR devices that show excellent Q and short term stability. The 160 MHz devices have Q's that range up to 48,000, series resonance impedance less than 40 ohms, and short term stability that approaches 5 parts in 10^{11} for 1 second samples. The temperature performance of these devices is essentially that of ST-cut quartz,⁶ except that room temperature turnover occurs at an orientation of 39.3°.

The etching process used is a reactive fluorine plasma process capable of etching quartz at 3000Å/min.⁷ Etching grooved resonators 0.05 wavelengths deep (about 8500-9000Å) typically requires about 30 minutes. Fabrication techniques developed are simple, straightforward, and easily adapted to large scale batch processing.

The major limitation is that long term stability is no better than 1 part in 10^8 per day.

Fabrication

Substrate

Figure 1 gives a general overview of the one-port SWR device used. Each grating region consists of 275 grooves. The IDT is a 49 section transducer with cosine weighted apodization. The total aperture is 60 wavelengths and the grating-IDT separation is about 5 wavelengths. All units were fabricated on crystalline quartz.

Quartz SWR's have been made at various orientations. Figure 2 shows the relationship between orientation angle and the turnover temperature. The results reported in this paper are measured primarily on devices oriented at 35.821° to permit oven operation at turnover during measurement.

The substrates are 15 mm. in diameter to take advantage of existing processes, fixturing, and mounts

for bulk wave devices. The thickness is nominally 1.25 mm. The substrate is x-ray oriented to $\pm 15^\circ$ of arc, lapped, suitably polished and cleaned for vacuum plating.

Since the device will be eventually vacuum-brazed to a mount, the edges are vacuum-plated with 100Å of chrome followed by 5000Å of gold. After a change of evaporation masks, the units undergo a vacuum bake at 10^{-7} torr then receive 1200 \pm 100Å of aluminum in the active region. Immediately after removal from vacuum, photoresist is spun on, baked, exposed, developed, and the unwanted aluminum is etched away. The photomask used is a single mask with gratings and IDT properly spaced.

Etching

Figure 3 shows a device ready for etching. The aluminum pattern from the previous steps acts as an etch resist. The process used removes quartz about 10 times faster than it does aluminum. In order to preserve the advantages of a single metallization and single photoresist exposure, it is necessary to protect the IDT region during etching. Figure 4 shows a nylon mask that is placed over the IDT during etching. This has an additional advantage in that it also partially shadows the first few grating lines. During etching, a rather smooth transition occurs from the unetched surface to the grating because of this shadowing.⁸

The system used is shown in Figure 5. This is a reactive fluorine plasma system used to etch quartz and other materials extremely rapidly.⁹ As reported previously, etch rates of 3000Å per minute are possible.⁷ However, as this high of an etch rate destroys the aluminum etch resist too fast, the rate is usually decreased to about 300Å per minute. With this etch rate, the aluminum etch resist has been almost removed when the proper depth is reached, eliminating the necessity of doing this later.

During the etching process, the resonant frequency can be monitored. Figure 6 shows the relationship between depth of etch and the change in resonant frequency. For our process, Li's "B" factor³ has been experimentally determined to be 22.5. By monitoring during etching, the final device frequency on single units has been trimmed to within ± 500 Hz. of the desired frequency.

While determining the correct value for "B", we observed, using network analyzer, that the total insertion loss continued to decrease and the device Q continued to increase at depths far beyond that reported by Li as possibly causing bulk mode scattering at the groove interface.¹⁰ Continued measurements at 160 MHz and 320 MHz indicate that the total insertion loss continues to decrease (Q continues to increase) up to groove depths of 0.04 - 0.05 wavelengths. Beyond

this, the insertion loss and Q tend to flatten out, and then, beyond about 0.07 wavelengths, the insertion loss starts to increase, and Q decreases. Occasionally in this region spurious mode responses are also seen.

As a result, our current practice is to design for a resonant frequency at a groove depth of 0.045 wavelengths. Final adjustment to the correct resonant frequency can then be made by adjusting the actual etch depth while being assured that the insertion loss is at a minimum and Q is at a maximum.

Mount

After the etch process, the substrates are then vacuum-brazed onto a ceramic header developed for 5 MHz bulk wave devices. The brazing material is a gold-germanium alloy that melts at 356°C. After brazing, a copper top-cap is cold-welded to the header. The units are then placed on a bakeout head, connected by a tubulation previously brazed to the top-cap. All units then undergo a vacuum bake at 10^{-7} torr, at 200°C for 8 to 48 hours. This process is identical to that used for high precision bulk-mode resonators that show long term aging approaching parts in 10^{11} per day. After baking, the tubulation is cold-welded shut by pinching-off at a minimum pressure of 1×10^{-7} torr.

Measurement

After preliminary screening, eight units were selected and installed in oscillators. The oscillator is a rather conventional single transistor unit designed for a previous instrument project. Figure 7 shows an aluminum oven which was constructed to hold the eight resonators and oscillators. After preliminary adjustments, the oven was set at the nominal turnover temperature of the resonators.

Resistance and Q

Preliminary screening required the measurement of the series resonance resistance and the Q of the resonator. The measurement setup¹¹ is shown in Figure 8. Q was measured using the phase-slope method at resonance. Resistance was determined from the measured voltages and phase angles, after making appropriate correction for known stray reactances. A summary of the measured data is shown in Table I. One resonator, 11 B 4, had been opened to air accidentally prior to making the measurements.

Table I shows Q values ranging from 26,000 to 48,000 with the majority of the units in excess of 38,000. All of these units were etched at least 0.05 wavelengths deep at 160 MHz. From the known material losses of quartz, the maximum achievable Q at this frequency should be 65,000. In practice, the units discussed here have a large, complicated IDT structure in the acoustic cavity which is known to limit the obtainable Q. Experimental Q's approaching 80% of the theoretical maximum indicate that the acoustic losses in such a device must be extremely low. As any conversion to bulk waves at a groove interface would represent a significant acoustic loss, one must conclude that for devices made on quartz using the described methods, conversion to bulk waves must be virtually nonexistent.

Preliminary screening also required the measurement of the turnover temperatures of the devices, and the determination of the frequency-tempera-

ture coefficients of the devices. Although not cut at the ST orientation of 42.75°, these units show the same temperature dependence with a second-order temperature coefficient at turnover of -36 parts in 10^3 per degree C squared. The final selection of units was determined by picking four from each experimental group that were the closest in turnover temperature. The mean of the eight units was 54.15°C.

Long Term Aging

After installing the eight resonators and oscillators in the oven, daily measurements of the resonant frequency were taken for approximately six weeks. During the entire test, the oscillators were always working, the oven temperature was controlled to within $\pm 0.01^\circ\text{C}$, and the system monitored using a mechanical scanner and a diode switching matrix to connect the oscillator outputs to appropriate counters. The system time base was a Model 107 BR Quartz Frequency Standard that has been in continuous operation for a number of years and is known to be stable to better than one part in 10^{11} per day.

After the first three to six days, the units settled down and aged at essentially a constant rate for the remainder of the testing time. Table I also contains a summary of the final long term aging rate for the eight test units. The highest aging rate belongs to the unit which has been opened to air. The cause for the negative aging rate of unit 12 B 6 is unknown, but may be due to a resin leak sealant used on several cans suspected of being leaky while on the bakeout head.

The experimental data is in rather close agreement with that reported elsewhere. As all processes used were identical with those that produce bulk mode units that age at a rate almost three orders of magnitude better than the SWR's, the results are disappointing. The test was terminated rather suddenly when electrical power was cut to allow construction of a new lab facility.

Short Term Stability

Fig. 9 shows the measurement setup used to determine the short term stability. All measurements were made after approximately 5 weeks of aging. Measurements were performed only in the time domain since this was compatible with the equipment available. A summary of the results obtained for 1 second sampling times is also shown in Table I. In all cases, $\Delta f = 1000$ Hz, and the number of samples was 200.

Fig. 10 shows the relationship between the short term stability, expressed as the Allen Variance, and the sampling time for one of the eight units. The shape of the curve for all of the other units was identical.

Table I, upon closer inspection, shows some interesting results. There is a high correlation between the short term stability measurements and the resonator Q, as expected. In addition, the group marked "11 B" had significantly better short term stability than the "12 B" group. A check of the previous history of these substrates showed that the "11 B" group was all reprocessed units. These substrates had previously been made into SWR's, and for one reason or another were relapped and repolished. A possible conclusion is that group "12 B" had not been sufficiently lapped before polishing, as both groups were polished to the same standards.

Drive Level Dependence

The test setup shown in Fig. 8 was also used to determine the drive level dependence of the SWR. The resonant frequency was measured as the drive level was changed from -60 dBm to 0 dBm. The results are shown in Fig. 11. Drive levels below -35 dBm have no effect on the resonant frequency. The region between -20 and -35 dBm is probably due, in part, to elastic nonlinearities. For drive levels greater than -20 dBm, the rapid frequency change is probably due to thermal effects. The unit was tested at room temperature, although its turnover temperature was about 55°C.

Conclusion

The experimental results shown here indicate that with this process, one can etch considerably deeper than 0.02 wavelengths and still observe no detectable bulk wave scattering in quartz. Q's were achieved at 80% of that theoretically possible, even with a large IDT in the acoustic cavity. Further, the excellent short term stability of these devices make them promising candidates for use as high frequency oscillator crystals. Unfortunately, their usefulness, at the present time, is limited by the rather poor long term performance.

Acknowledgements

The authors would like to acknowledge the efforts of Henry Yoshida, George Nelson, and Johnny Ratcliff in the fabrication and polishing of the substrates.

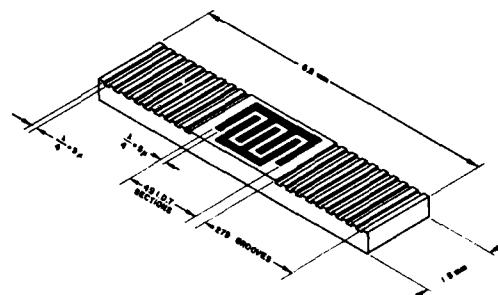
References

- ¹ E. J. Staples, "UHF Surface Acoustic Wave Resonators," *Proc. 28th Ann. Symp. on Frequency Control*, pp. 280-5, 1974.
- ² K. M. Lakin, T. Joseph, and D. Penunuri, "Planar Surface Acoustic Wave Resonators," *Proc. 1974 Ultrasonics Symposium*, pp. 263-267, 1974.
- ³ R.C.M. Li, J. A. Alusow, and R. C. Williamson, "Surface-Wave Resonators Using Grooved Reflectors," *Proc. 29th Ann. Symp. on Frequency Control*, pp. 167-76, 1975.
- ⁴ S. P. Miller, R. E. Stigall, and W. R. Shreve, "Plasma Etched Quartz SAW Resonators," *Proc. 1975 Ultrasonics Symposium*, pp. 474-7, 1975.
- ⁵ P. Hartemann, "Acoustic Surface Wave Resonators Using Implanted Gratings," *Proc. 1975 Ultrasonics Symposium*, pp. 303-6, 1975.
- ⁶ M. B. Schulz, B. J. Matsinger, and M. G. Holland, "Temperature Dependence of Surface Acoustic Wave Velocity on α -Quartz," *Proc. IEEE*, pp. 1361-2, Sept. 1970.
- ⁷ J. A. Kusters, C. A. Adams, H. E. Karrer, and R. W. Ward, "Analytical and Experimental Investigations of 32 KHz Quartz Tuning Forks," *Proc. 30th Ann. Symp. on Frequency Control*, pp. 175-83, 1976.
- ⁸ The effect of shadowing was discussed in: J. A. Alusow, R. C. M. Li, and R. C. Williamson, Paper P4, "Suppression of Bulk-Wave Scattering Loss in SAW Resonators," presented at the 1976 Ultrasonics Symposium.

⁹ U. S. Patent 3,971,684, July 27, 1976.

¹⁰ R. C. M. Li, J. A. Alusow, and R. C. Williamson, "Experimental Exploration of the Limits of Achievable Q of Grooved Surface-Wave Resonators," *Proc. 1975 Ultrasonics Symposium*, pp. 279-83, 1975.

¹¹ First discussed in: C. A. Adams, J. A. Kusters, and A. Benjaminson, "Measurement Techniques for Quartz Crystals," *Proc. 22nd Ann. Symp. on Frequency Control*, pp. 248-58, 1968.



SURFACE ACOUSTIC WAVE RESONATOR

FIGURE 1

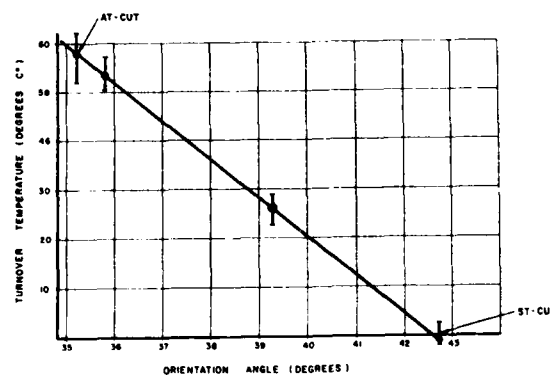


FIGURE 2

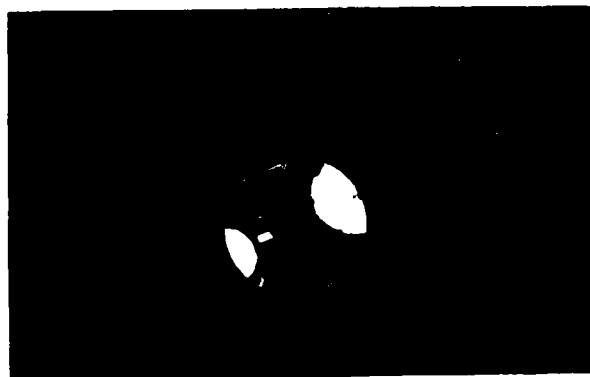


FIGURE 3 - SWR Ready for Etching.

CRYSTAL NO	R_0	Q_0	SHORT TERM STABILITY %	LONG TERM AGING
11 B 4	44.5	31,126	9.0×10^{-11}	4.1×10^{-8} day
11 B 5	35.7	43,008	6.0	1.9
11 B 6	23.3	48,201	5.1	1.3
11 B 7	28.8	38,352	7.0	2.3
12 B 4	24.5	36,996	15	2.4
12 B 6	23.7	46,703	15	-2.7
12 B 7	37.3	26,768	41	2.2
12 B 8	25.9	38,451	16	1.1

* 1 sec SAMPLING TIMES, $\Delta F = 1000$, $M = 200$

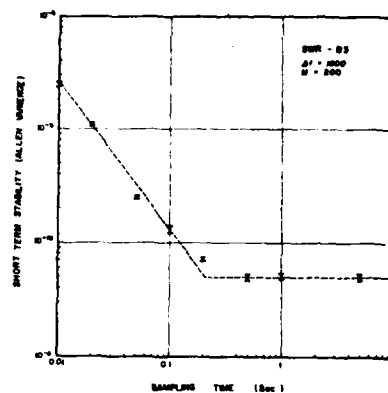


TABLE I

FIGURE 10 - Short Term Stability.

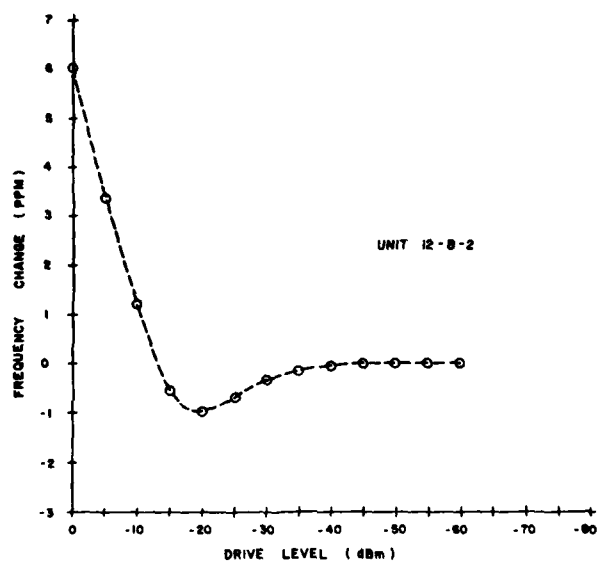


FIGURE 11 - Drive Level Frequency Effect

PIEZOELECTRIC SHEAR SURFACE WAVE GRATING RESONATORS

B. A. Auld and Bing-Hui Yeh
Edward L. Ginzton Laboratory
W. W. Hansen Laboratories of Physics
Stanford University
Stanford, California

Summary

This paper describes a new type of surface acoustic wave grating resonator in which the particle displacement of the surface wave is parallel to the surface. By contrast, the now well-known SAW (Rayleigh wave) grating resonator has its particle displacement in the sagittal plane.

In the SAW resonator the function of the grating structure, which consists of either grooves or metal strips, is to provide two highly reflecting Rayleigh wave mirrors, between which a standing wave is excited by means of an interdigital transducer. The basic function of the grating in the horizontal shear (SH) type of surface wave resonator considered here is quite different. An SH surface wave cannot exist on a homogeneous (unlayered) substrate in the absence of some periodic variation, such as a grating, along the surface. That is to say, the surface wave in this case is a vibrational mode of the grating itself. On the other hand, the grating is now not required to realize a highly reflecting mirror, because the SH motion reflects without spurious mode coupling at a traction-free boundary placed in any symmetry plane of the structure. For this reason this new type of resonator promises a substantial advantage in miniaturization compared with the conventional SAW resonator.

SH surface wave resonators on PZT-8, Y-cut X-propagating LiNbO_3 , and ST quartz have been fabricated and tested. Excitation was by means of an interdigital structure deposited on top of the grating teeth, and the dimensions were chosen to give a resonance in the region of 1 to 2 MHz. The groove depth was in the range of 0.01", and it was found that the diamond saw fabrication technique used did not provide adequate precision. Consequently the quality factors realized were low (< 3000), and use of relatively shallower etched grooves at higher frequencies is clearly called for.

Since a sufficient condition for the existence of this type of surface wave is a periodicity of the conditions along the surface, another technique for trapping the wave at the surface is deposition of an array of metal strips.

Key words Resonator, Grating, Horizontal Shear, Interdigital Transducer, PZT, Lithium Niobate, Quartz.

Introduction

In an earlier paper¹ the existence of a horizontally polarized shear (SH) surface wave on a corrugated substrate was demonstrated by virtue of the exact analogy between this elastic wave problem and the corresponding electromagnetic problem. An extensive literature exists for the latter case and the solutions given were found to be in good agreement with experimental results obtained for shear surface waves on a corrugated aluminum substrate. Observations were made by fabricating a finite length of corrugated surface (or grating) and measuring the transmission resonances with thickness shear transducers bonded to the ends of the finite length of grating. The surface nature of the wave was

confirmed by measuring gratings on substrates of different depths.

Although previous experiments confirmed the applicability of the electromagnetic solutions to the elastic wave problem, the transducers used excited both the surface wave of the grating and thickness waves of the substrate. The mode spectrum was therefore cluttered with many spurious responses and entirely unsuitable for resonator applications. The present paper describes an investigation of SH surface wave resonances on piezoelectric substrates. Efficient selective excitation of the grating is then achieved by suitably choosing the substrate orientation and depositing a transducer electrode on top of each tooth. As will be seen, individual resonant modes of a finite length of grating may be excited by suitably choosing the distribution of voltages applied to the electrode array.

It should be pointed out that the SH surface wave considered here is intimately related to the surface-skimming shear wave.^{2,3} The latter wave or, more properly, radiation pattern consists of a horizontally polarized shear elastic vibration skimming along the surface and slowly diffracting into the substrate. Diffraction losses are determined by the vertical directivity of the interdigital transducer (IDT) used for excitation. Consistent with the analogy of the IDT as an end-fire antenna array, the radiation pattern is sharpened and diffraction losses are reduced by increasing the length of the IDT. Addition of a grating structure to the surface permits the existence of a genuinely bound SH wave, which travels at a velocity slower than that of a bulk SH wave. If the IDT is suitably designed to synchronize with this bound wave, as in the case of a Rayleigh wave transducer, there will be essentially no diffraction loss into the substrate.

SH Waves on an Infinite Grating

One way of picturing an SH grating vibration is to imagine it as evolving from the standard tuning fork resonator shown in the upper left of Fig. 1. An analogous type of tuning fork, in which the arms move in face shear, is shown on the right. Stacking of a number of these resonators in an array leads to the tuning fork grating shown at the bottom of the figure, in which the dashed lines are traction-free surfaces. The basic SH grating (Fig. 2) evolves from this as the dimension is extended to infinity along the particle displacement direction and the individual supports are replaced by a continuous substrate. With a fixed tooth spacing d , the frequency increases with decreasing length of the teeth, just as in the case of the original tuning fork. The grating configuration provides a means for realizing a tuning fork type of resonance at frequencies where a single fork becomes too small to fabricate and mount.

Because the spatial period of the vibration in Fig. 2 is $2d$ the displacement field in the substrate can be written as the Fourier series shown in the figure, where a_n is the amplitude of the n^{th} Fourier component and the γ_n exponential coefficient γ_n is

related to the wavelength of a bulk shear wave by the equation on the line below. It follows from this that γ_n is real when $2d$ is less than a bulk shear wavelength. In this case all of the Fourier components decay exponentially into the substrate - that is, the motion is a surface vibration bound to the grating. This argument does not, of course, prove the existence of such a vibration, but its existence has already been demonstrated analytically in the case of the analogous electromagnetic problem.^{4,5}

The vibration shown in Fig. 2, which has a phase shift of π from one grating tooth to the next (the π -mode), is only one of many that can exist on this periodic structure. In the previous discussion, the grating was regarded as essentially an infinite array of tuning forks. Alternatively, one may look at the teeth as an array of cantilever supported face shear plate vibrators that are lightly coupled, one to the next, through the substrate. The vibration spectrum consists of a continuous distribution of coupled modes, analogous to the modes of a periodically mass-loaded vibrating string. In this case the phase shift from section to section is related to a continuous wave number $k = 2\pi/\lambda$, which takes the value π/d for the π -mode discussed above (Fig. 3).

As shown in References 4 and 5 the relationship between ω and k for this grating surface wave has the same form as for waves on the periodically loaded string. The frequency of the π -mode ($k = \pi/d$ in Fig. 3) corresponds to the lower edge of the stop band. Above this frequency the surface wave is nonpropagating (or cut-off). As the depth of the grating grooves is decreased, the frequency of the π -mode increases until the V_{SHEAR} line is reached. This corresponds to the surface-skimming shear wave discussed above. In the so-called slow wave region below this line the solution is always a surface wave.

It should be emphasized that very little slowing is required to produce a well-confined surface wave. A Rayleigh wave, for example, has a phase velocity that is only some five percent below the bulk shear velocity but is confined to a depth less than a shear wavelength. One needs, therefore, only a shallow grating to trap the SH wave on the surface.

Finite Grating Resonators

To produce a standing surface wave resonance the grating structure must be terminated in a pair of mirror reflectors. In the standard SAW resonator these mirrors are realized by long (several hundred periods) grating arrays designed to operate in the cut-off region. For the SH surface wave resonator this is not necessary. A mirror can be realized by terminating the grating in a suitably located traction-free boundary. In the case of the π -mode this is easily seen by examining Fig. 2, where the particle displacement is along x and varies with y and z . From the symmetry of the vibration one has that the displacement u_x is maximum with respect to the z variation at the plane denoted by a dashed line in the figure. This means that the strain component S_{xz} and the stress component T_{xz} are zero on this plane. Since u_x is a function only of y and z , the stress components T_{yz} and T_{zz} are also zero - just the conditions required for a traction-free boundary, which acts as a perfect mirror. By further symmetry arguments one can show that the same boundary conditions acts as a perfect mirror for a surface wave with any wavenumber k .

Figure 4 gives the profile of an N section resonator contained between two such mirror reflectors.

As in any standing wave resonance, the resonance condition is that the length L be integral number n of half wavelengths or, equivalently, that $k = n\pi/L$. For a 10-section resonator there are therefore ten modes of resonance, with frequencies determined from the dispersion diagram by the construction shown on the figure. A particular mode may be excited by applying the corresponding distribution of voltages to the electrodes located on the tops of the teeth. Our experiments have been performed on the π -mode in which alternate electrodes are excited 180° out of phase, as in a conventional IDT.

We have fabricated and tested three SH grating resonators with groove profile dimensions as given in the upper left of Fig. 4 and resonant frequencies for the π -mode in the range of 1 to 2.5 MHz. The resonator proper is defined by an electroded region on the top surface of a grooved block, large enough to eliminate edge effects and to permit probing of the vibration pattern outside the electrode region by means of small rubber damping pads. The grooves were first cut with a diamond saw, the top surface and reflecting edges were then polished, the electrodes deposited and the gold wire leads attached.

In these initial experiments no attempt was made to polish the inside of the grooves. As will be seen, this leads to problems with resonance broadening and coupling into spurious modes due to grating nonuniformity and surface roughness. It is clear that the best way to make these structures is by deep etching techniques.⁶

The importance of groove depth uniformity is clear from the dispersion curves in Fig. 4, where it is seen that the frequency of the π -mode is strongly dependent on the groove depth h . Nonuniform groove depth therefore causes different parts of the grating to resonate at different frequencies. This effect was observed in some of our gratings, where the vibration was found by mechanical probing to be localized in a small region of the grating. Tighter tolerances on the fabrication procedure were found to reduce this effect. It appears from these results that the resonance could be confined to a desired region of the surface by deliberately tailoring the depth profile of the grooves.

Mechanical probing of the resonators with small rubber pads confirmed the surface wave nature of the vibration and also demonstrated a lateral confinement of the vibration to the electroded region. A lateral decay distance of the order of 1 cm was observed outside the electroded region. This is due to mass loading by the electrodes, which effectively increases the depth of the grooves under the electrodes. Because of this lateral confinement transverse modes are also observed, as in standard SAW resonators.

PZT-8 Ceramic Resonator

The insert of Fig. 5 shows a highly schematic representation of the resonator geometry. Poling is in the direction indicated by the heavy arrow. Since it is not possible to pole over a 3" length, the block was fabricated from six 1/2" pieces carefully ground and bonded together with epoxy. No repolishing was performed after sawing the grooves, and the groove depth was measured to be approximately five percent greater at the left end of the grating than at the right. When the two halves of the grating were excited independently, different resonant frequencies were obtained - as expected, the lower frequency corresponding to the larger groove depth.

Figure 5 shows the measured input impedance characteristics for the entire grating - the small peak at 1.120 MHz corresponding to the deep end of the grating. A variation of groove depth in the order of five percent corresponds to a frequency variation in the order of one percent. The estimated resonator Q is of order 200, compared with a material Q of order 400 determined by pulse echo measurements in a large block.

We have also tested the same geometry as a Rayleigh wave grating resonator, with the poling in the vertical direction in Fig. 5. As anticipated the Q is much lower (of order 50) because traction-free boundaries do not act as good mirrors for Rayleigh wave motion.

Y-X Lithium Niobate Resonator

A lithium niobate resonator was made in order to more clearly isolate the effects of surface roughness and resonator geometry on the Q-factor. As shown in Fig. 6 a resonator of somewhat different dimensions was made on a single crystal block, using the same fabrication method. The top surface is Y-oriented and the grooves are along Z. By cutting the grooves before polishing the top surface and finishing afterwards, grooves with clean upper corners were obtained. They were, however, very fragile and had to be handled with great care.

The impedance curve in the figure is for an unsupported sample and exhibits a very dense spectrum of spurious bulk modes in the region of the main and transverse surface wave resonances. Appearance of strong spurious modes only in the region of surface wave resonance indicates that the spurious coupling is through mechanical imperfections of the grating and not directly from the electrodes.

Figure 7 shows, on the same scale, the impedance characteristics after glueing the ends of the lithium niobate block. The bulk modes are now strongly suppressed, although some spurious is still apparent near to two transverse resonances on the high frequency side of the main resonance. There is now a strong surface wave resonance with a maximum impedance of 300 k Ω , and Figs. 8 and 9 show that the maximum spurious response over the range from 1 to 7.5 MHz is 4 k Ω .

Figure 10 gives on a logarithmic scale the detailed impedance response in the vicinity of the resonance and antiresonance points. Because of the large number of spurious modes near the series resonance point, it is not possible to calculate the Q by the standard procedure, but it is estimated to be not more than 2000. This very low value is clearly due to the energy loss coupled into a large number of bulk modes and subsequently dissipated in the supports. Further precision in grating fabrication is obviously called for.

ST Quartz Resonator

Figure 11 shows the geometry of a grating resonator on a 39.3° "ST" quartz plate with the grooves along the X direction. Note that the thickness of the substrate is much less than in the other examples. Although the resonator is unmounted the spurious mode response is small compared with the lithium niobate case, and the response is clean outside the 10 KHz frequency range shown.

As in the PZT resonator the low frequency peak is attributed to nonuniformity of groove depth. Since this second peak occurs very close to the main series resonance point, it is not possible to arrive at an ac-

curate evaluation of Q. We estimate a value in the order of 3000.

Conclusions

In summary, we have experimentally verified the surface character of SH vibrations on deep grating structures and have measured some of the properties of grating resonators of this type operating on PZT-8, YX lithium niobate and ST quartz in the frequency range of 1 to 2 MHz. Q-factors obtained are low, not more than 3000. This is due in part to technical difficulties in accurately and uniformly fabricating the large grooves required at these low frequencies and in part to mode scattering due to the effect of substrate anisotropy on the behavior of the traction-free reflector surfaces.

The fabrication problems encountered point up the need to study shallower grating resonators operating at higher frequencies. As pointed out earlier, only a small amount of slowing is needed to trap the SH vibration on the surface, and the grooves need only be a small fraction of a wavelength in depth. Since the periodicity needed for wave trapping may also be induced by periodic boundary conditions on the surface, an attractive alternative is mass-loading or electrical short circuit strips deposited on the substrate surface.⁷

The major potential advantage of the SH grating resonator over the standard SAW structure is in its small size. Because the grating itself does not serve as a mirror, only a small number of periods is required. Also, the presence of a spectrum of resonator modes that can be selected by appropriate coding of the applied electrode voltages suggests the possibility of small multipole monolithic filters at very high frequencies.

Acknowledgements

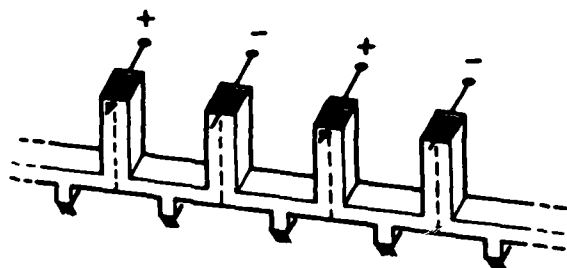
The authors wish to acknowledge the able technical assistance of W. Bond, L. Goddard, G. Kotler and D. Walsh.

This work was sponsored by the Joint Services Electronics Program under Contract N00014-75-C-0632.

References

1. B. A. Auld, J. J. Gagnepain and M. Tan, *Electronics Letters* **12**, pp. 650-651 (1976).
2. T. I. Browning and M. F. Lewis, these Proceedings.
3. K. H. Yen, K. L. Wang and R. S. Kagiwada, these Proceedings.
4. R. A. Hurd, *Can. J. Phys.* **32**, pp. 727-734 (1954).
5. R. A. Collin, "Field Theory of Guided Waves," pp. 465-469, McGraw-Hill (1960).
6. C. A. Adams and J. A. Kusters, these Proceedings.
7. A. A. Oliner and A. Hessel *Trans. IRE AP-7*, pp. 5201-5208 (1959).

TUNING FORK "GRATING"



TUNING FORK RESONATORS

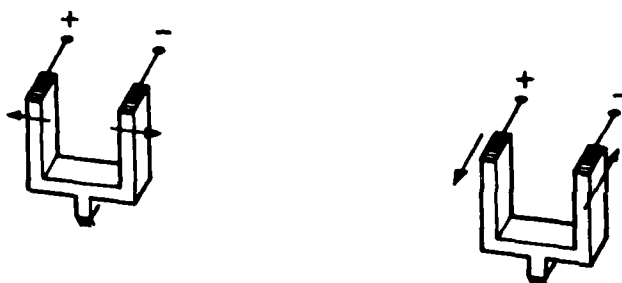
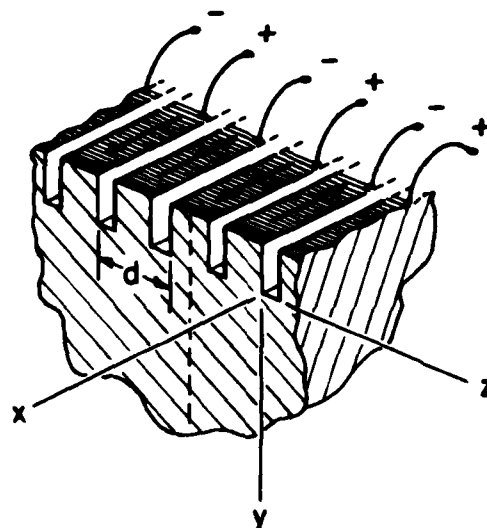


FIGURE 1. A Conventional flexural tuning fork resonator is shown in the upper left and a "face-shear" analogue in the upper right. The bottom of the figure shows a primitive horizontal shear grating resonator realized by stacking "face-shear" tuning forks.

INFINITE GRATING ON AN ISOTROPIC SUBSTRATE



$$u_x(y, z) = \sum_n a_n e^{\gamma_n y} e^{-i \frac{n\pi}{d} z}$$

$$-\gamma_n^2 = \left(\frac{2\pi}{\lambda_{\text{SHEAR}}} \right)^2 - \left(\frac{n\pi}{d} \right)^2 < 0$$

WHEN $d < \frac{\lambda_{\text{SHEAR}}}{2}$

↓
SURFACE VIBRATION

FIGURE 2. Horizontal shear grating on a semi-infinite substrate. Vibration is in the π -mode.

WAVES ON PERIODIC STRUCTURES

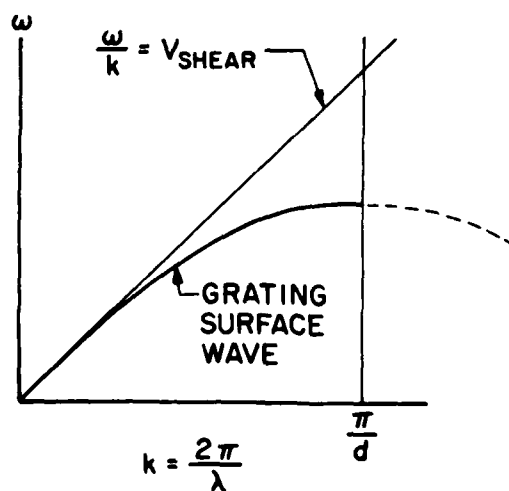
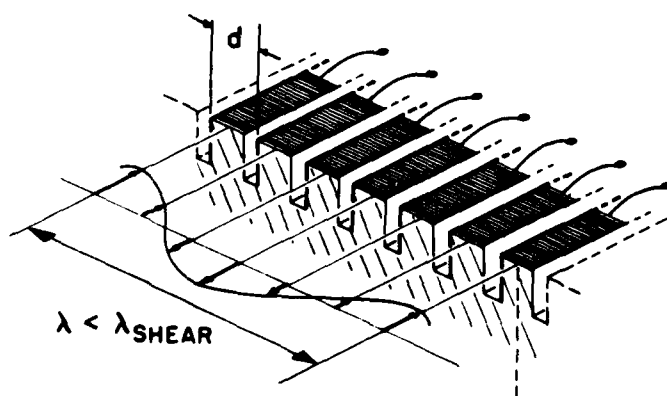


FIGURE 3. Typical form of the lowest branch of the dispersion relation for the horizontal shear grating.

RESONATOR GEOMETRY

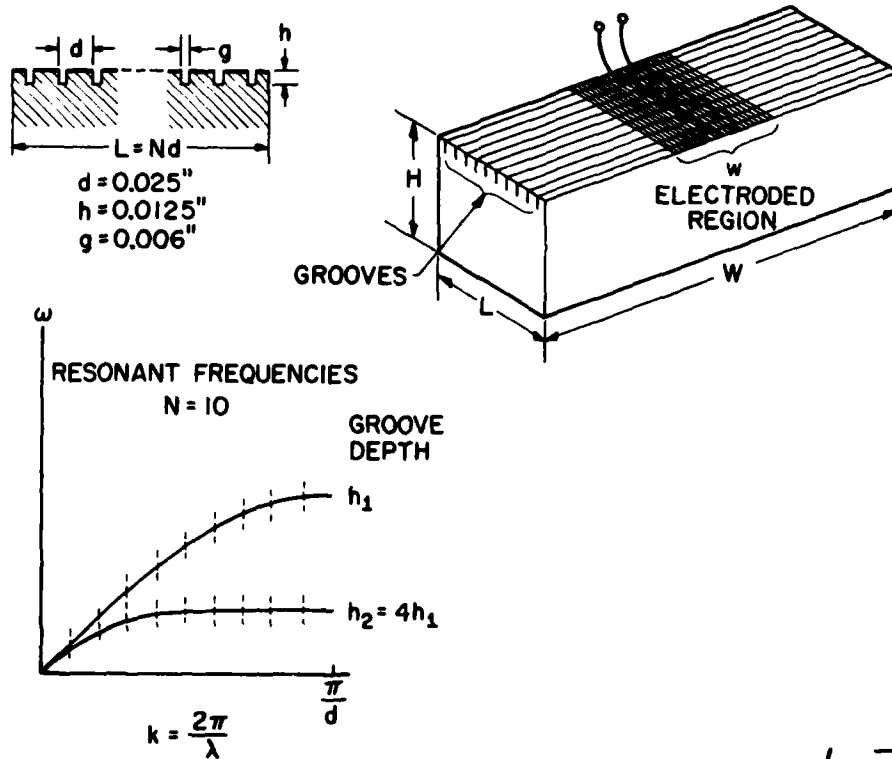


Fig. 4 Basic resonator structure and relationship between the dispersion relation and the resonant mode spectrum.

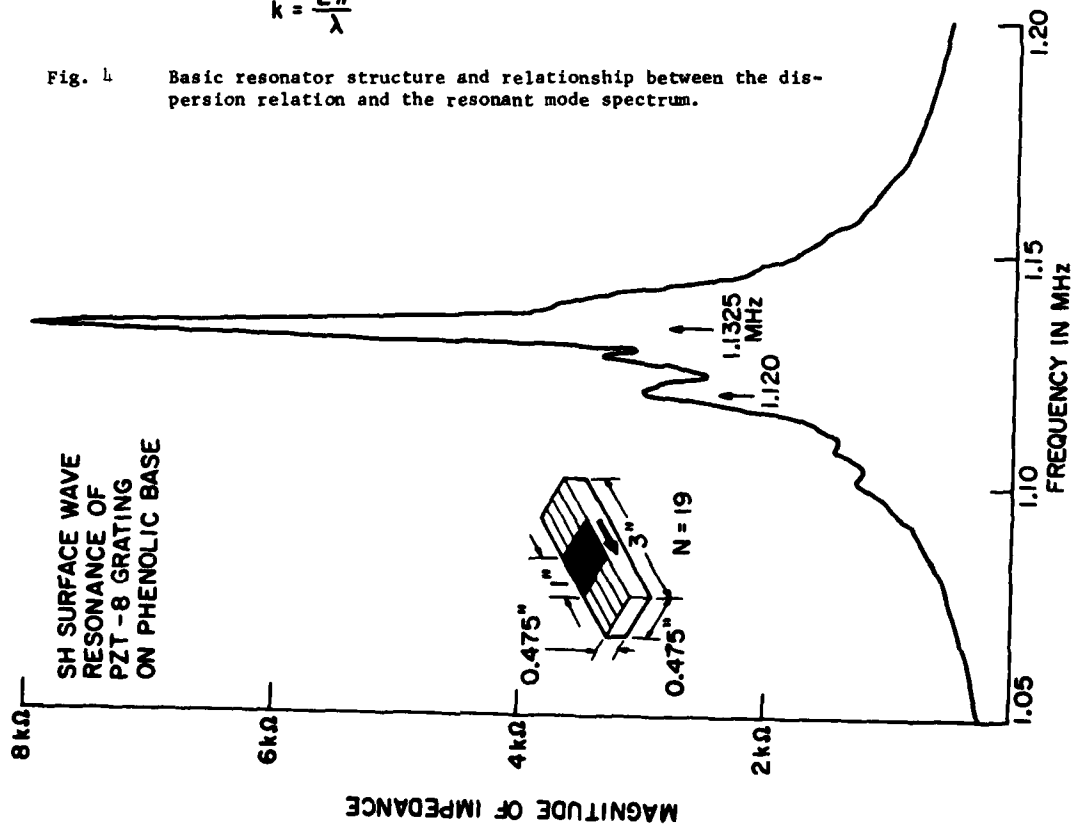


Fig. 5 Measured impedance-frequency curve of PZT-8 resonator. Poling direction is indicated by the heavy arrow.

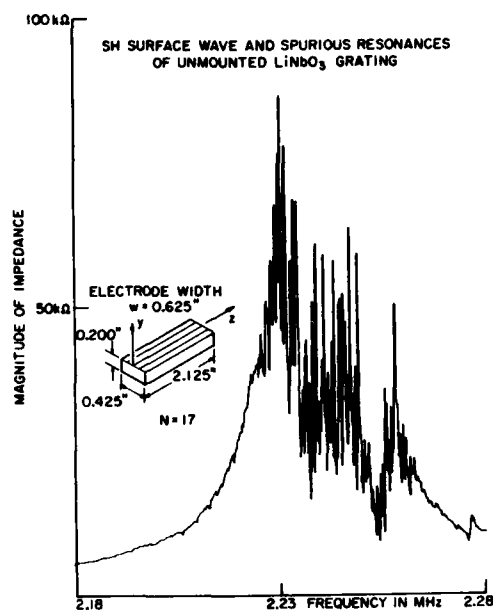


FIGURE 6. Impedance-frequency response of Y-oriented X-propagating LiNbO_3 grating resonator, showing a dense spectrum of spurious bulk resonances superposed on the main and trans-surface wave resonances. Area of the electrode region (not shown) is $0.425'' \times 0.625''$.

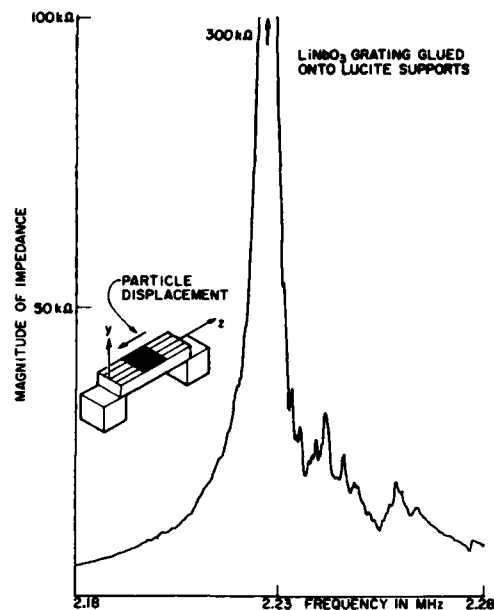


FIGURE 7. Same as Fig. 6, but with ends of the substrate glued to lucite supports. Note the two transverse mode resonances, with superimposed spurious resonances, on the high frequency side of the main resonance.

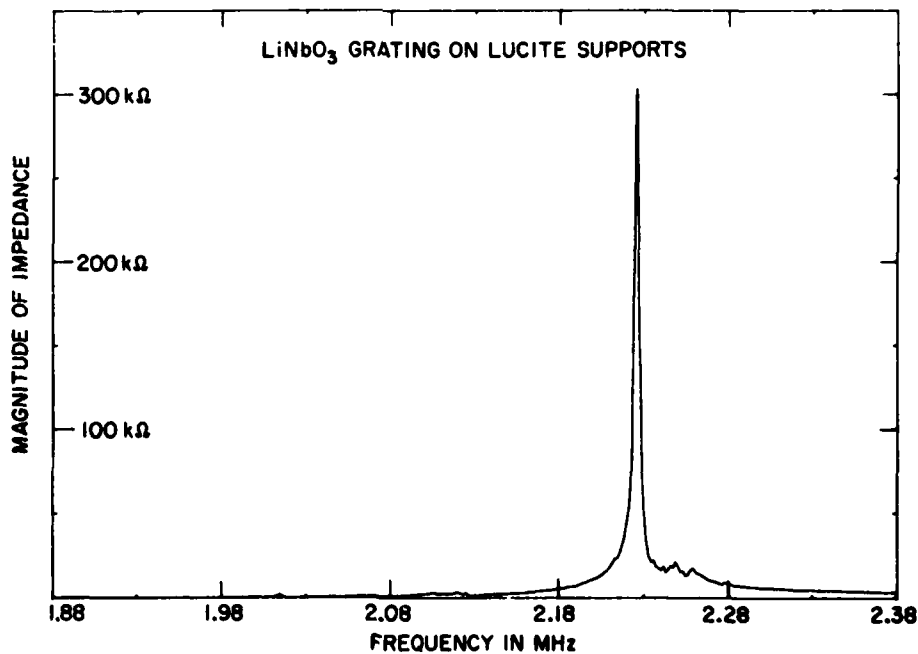


Fig. 8 Impedance-frequency curve of the LiNbO_3 grating, showing the peak of the main resonance and a wider portion of the skirts.

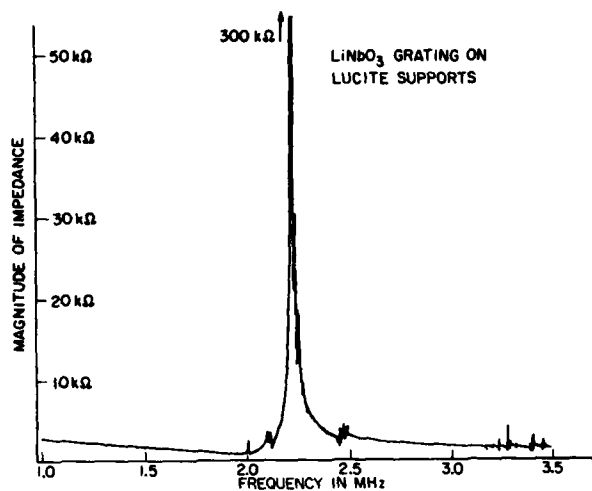


FIGURE 9. Illustration of the spurious mode response between 1 MHz and 3.5 MHz. Measurements out to 7.5 MHz showed a maximum spurious impedance peak of 4 kΩ over this frequency range.

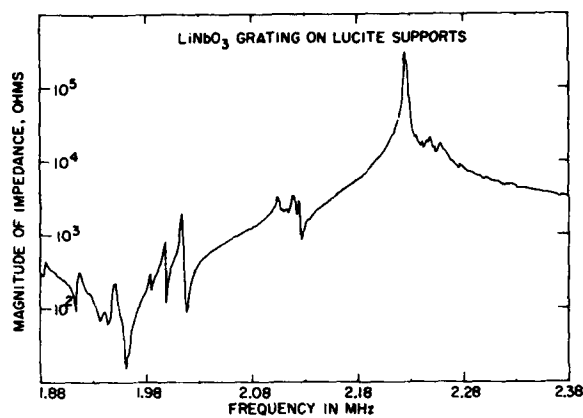


FIGURE 10. Logarithmic impedance curve in the vicinity of the resonance and antiresonance points.

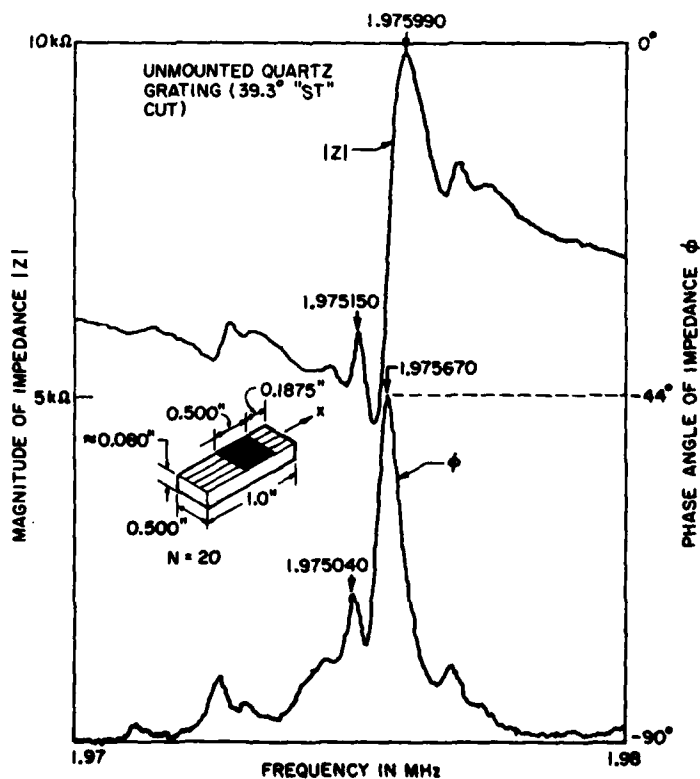


Fig. 11 Impedance-frequency curve of unmounted quartz grating on a 39.3° "ST" plate.

A NEW CLASS OF QUARTZ CRYSTAL OSCILLATOR CONTROLLED BY

SURFACE-SKIMMING BULK WAVES

T I Browning and M F Lewis
RSRE, Malvern, Worcs, UK

Summary

The attractions of surface acoustic wave (SAW) components for use in oscillators and frequency filters have been reported in various papers presented at the Annual Frequency Control Symposia since 1974. The present paper describes an important development in this field whereby the orientation of the substrate is so chosen that the input interdigital transducer (I.D.T.) launches a bulk wave which skims along the surface and is received by the output transducer long before it reaches any other surface of the substrate. This new class of component therefore superficially resembles the SAW component, and retains many of its attractions, e.g. planar construction and design flexibility. In addition, however, it possesses a number of advantages over SAW, especially in narrow band filters and oscillators. These include: (i) higher velocity and lower propagation losses, which enable operation up to about 1.6 times the highest SAW frequency on ST-quartz, (ii) potentially superior temperature coefficients, and (iii) insensitivity to surface contamination leading to lower long-term ageing.

In this paper we describe the conditions necessary for a substrate to support a surface-skimming bulk wave, together with details of the transduction and propagation. We show that the attainable insertion loss can be as low as 13dB (compared with 10dB for a typical SAW device), and that the frequency response is closely related to that of the corresponding SAW device. We also cover the important aspect of temperature coefficients of delay on quartz and other materials. The principles are illustrated by measurements on surface-skimming bulk wave (S.S.B.W.) filters and oscillators operating at frequencies up to 2.3 GHz.

Introduction

The interdigital transducer (I.D.T.) was first introduced by Mortley⁽¹⁾ for the transduction and reception of bulk acoustic waves travelling through the volume of a 3-dimensional solid sample. Such samples are frequently wedge-shaped, and their principal application is as dispersive delay lines for use in pulse compression radar. It was discovered by White and Voltmer⁽²⁾ in 1965 that the I.D.T. is also an efficient generator of surface acoustic waves (SAW) and this led to the development of a novel generation of planar acoustic wave devices, see for example, refs (3) and (4). From the foregoing it is hardly surprising that one of the principal problems encountered in SAW devices arises from the simultaneous generation by I.D.T. of unwanted bulk acoustic waves⁽⁵⁾. These bulk waves can interfere with the required SAW response in a number of ways, and are particularly troublesome in wideband filters. A recent review of the situation has been given by Milsom⁽⁶⁾.

Our own work has been principally in the field of SAW oscillators, in which the SAW element is essentially a narrow-band filter with prescribed phase and amplitude response. In such narrow-band SAW filters the effect of the bulk waves is often to introduce separate pass-band(s) of similar shape to the SAW passband, but at higher frequencies, directly reflecting the higher velocity of the bulk waves, see for example Fig. 1.

These passbands are caused by those bulk wave(s) which can propagate with k -vector parallel to the surface while still satisfying the surface boundary conditions. Such surface-skimming bulk waves (S.S.B.W.) travel with velocity and attenuation essentially unchanged from their values in the infinite medium. It is, of course possible to use one of these bulk wave responses rather than the SAW response to make a filter or to control the frequency of an oscillator⁽⁷⁾; this may be done, for example, by tuning out the static capacitance of the I.D.T.'s at one of the bulk wave frequencies, by damping the SAW response (section 1), or by choosing a substrate orientation for which the piezoelectric coupling (k^2) of SAW vanishes (section 5). Over the past few years we have studied S.S.B.W. propagation in various orientations of quartz, LiNbO₃ and LiTaO₃, and have developed a detailed understanding of many of the properties of S.S.B.W. devices including excitation, propagation, insertion loss, frequency response and temperature coefficient of frequency (or delay). The present paper presents the results of this study, and shows that S.S.B.W. retain many of the attractions of SAW devices while offering a unique combination of advantages over SAW. The paper is divided up as follows:

Section 1 describes some experiments illustrating the basic properties of S.S.B.W.

Section 2 discusses the effects of the stress-free boundary conditions on the propagation of S.S.B.W.

Section 3 discusses two theoretical approaches to the insertion loss of S.S.B.W. devices and shows experimentally that the loss can be as low as 13dB.

Section 4 is concerned with the frequency response of S.S.B.W. filters and its relationship to corresponding SAW filters.

Section 5 describes our search for orientations of piezoelectric crystals which are favourable for the propagation of S.S.B.W. and the construction of S.S.B.W. devices.

Section 6 describes miscellaneous other experiments on S.S.B.W. devices which are analogues of well-known SAW devices.

Section 7 concludes the paper.

1. Some basic properties of S.S.B.W.

A brief account of the present work has been reported previously⁽⁸⁾. To illustrate the properties of S.S.B.W., consider the frequency response of Fig. 1a. This is the measured response of a conventional SAW oscillator filter made on AT-quartz with propagation along the x -axis. The filter comprises two untuned I.D.T.'s, one with 100 finger pairs, the other with 70 finger pairs, each with 200 μ m aperture and with a centre-to-centre separation of 100 μ m (thereby satisfying the oscillator mode-selection requirement⁽⁹⁾). As well as the SAW response at 170 MHz, there are clearly visible two similar responses at 275 and

309 MHz, which frequencies are entirely consistent with the velocities of x-propagating SAW and bulk waves ($V_{SAW} = 3.15$, $V_{FAST SHEAR} = 5.1$ and

$V_{LONGITUDINAL} = 5.7$ km/s). Notice that the longitudinal wave is polarised in the surface, the fast shear wave is polarised within a few degrees of the surface, and that there is no detectable response from the slow shear wave (velocity 3.3 km/s) which is polarised (almost) perpendicular to the surface. These observations are true of all the cuts investigated to date, and are entirely consistent with the conclusions of Mitchell.⁽⁵⁾ When absorber is added to the 15λ gap between the I.D.T's the SAW response is reduced by > 30 dB, while the bulk wave responses are only reduced by a few decibels (Fig. 1b). This result strongly suggests that the acoustic propagation between the I.D.T's largely takes place below the surface of the substrate.

As discussed in section 5, a suitable substrate orientation for the observation of S.S.B.W. propagation is the AT-cut of quartz with propagation perpendicular to the x-axis; this cut has been the work-horse of our investigations. Fig. 2(a) shows the response of a "ladder-type" SAW oscillator filter on this cut of quartz, and contains the required S.S.B.W. response as well as a series of plate modes described by Wagers⁽¹⁰⁾. When the back face of the plate was ground to a small angle, i.e. rotated about the direction of propagation, the phase coherence of the plate modes was destroyed (Fig. 2b), but the S.S.B.W. response was unaffected. This experiment shows a clear distinction between the plate modes described by Wagers and our own surface-skimming bulk waves, which are beamed along the surface and received by the output transducer long before they reach the lower surface of the sample. A more detailed photograph of the S.S.B.W. response is given in Fig. 2(c), and it is clear that this response closely resembles the $\sin x/x$ response of the original SAW oscillator filter. We shall return to this aspect of S.S.B.W. devices in section 4.

2. Propagation of S.S.B.W.

In order to obtain some physical feeling for the results to be expected in real anisotropic (piezoelectric) crystals, let us first consider the case of an isotropic medium. For any direction of propagation (k - vector) the bulk waves which can propagate in the infinite medium are a pure longitudinal wave and two degenerate pure shear waves of arbitrary polarization in the plane perpendicular to k . We are seeking waves which, if propagating in an identical form in a finite substrate (in our case a plate), satisfy the stress-free boundary conditions on the principal surfaces of the plate. By analogy with the SAW problem⁽¹¹⁾ we shall assume that if such waves exist, they will propagate freely, and will form the basis of S.S.B.W. devices.

It is shown in appendix I that for a general isotropic medium, the surface boundary conditions can only be satisfied for a shear wave which is polarized in the surface; such a wave is frequently referred to as a horizontally-polarized shear wave. This result is closely related to the dispersion-free propagation of torsional waves in an isotropic cylinder⁽¹²⁾, recalling that in the torsional mode all particle motion is parallel to the surface.

When we consider general anisotropic crystals it is more difficult to find waves which satisfy the boundary conditions for two reasons:-

(i) the elastic constant matrix contains more non-vanishing elements, and

(ii) the anisotropy lifts the degeneracy of the shear waves and fixes their polarization directions so that, for an arbitrary substrate orientation with arbitrary direction of propagation, a horizontally polarized shear wave does not exist.

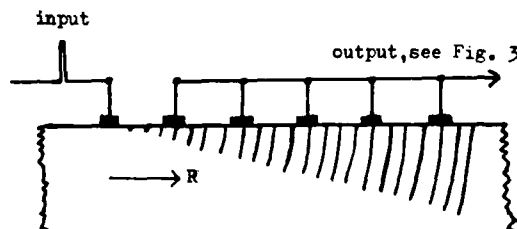
Nevertheless there is still an infinite set of substrate orientations/propagation directions which do support a horizontally polarized (quasi-) shear wave. To date we have only investigated the rotated Y-cuts in crystals with the elastic symmetry of quartz (i.e. including LiNbO_3 and LiTaO_3). In these cuts we find (appendix I) that the conditions for vanishing surface stresses are often identical to the conditions for the shear wave polarization to lie in the surface. This reinforces the conclusion reached from studying isotropic materials and shows that a good practical guide to finding suitable S.S.B.W. substrates is to seek orientations which support horizontally-polarized shear waves.

We need hardly add that there exist other waves which can travel a finite distance (say 100λ) with little attenuation even though they do not exactly satisfy the boundary conditions; an example is the longitudinal wave travelling on the x-axis of AT-quartz, Fig. 1.

Experimentally we have demonstrated the free propagation of S.S.B.W. on AT-quartz with propagation perpendicular to the x-axis. The input I.D.T. has an aperture of 300λ and contains only 7 finger-pairs, and so approximates to a line source of acoustic disturbance. The output I.D.T. is a "ladder" transducer containing 20 "rungs" each of 3 finger-pairs and spaced every 20λ (the closest rung being 20λ from the input I.D.T.). The input transducer is impulsed and the output waveform is shown in Fig. 3, which is the actual impulse response (voltage waveform) photographed on a wideband C.R.O. It is readily verified from this photograph that the output voltage goes as

$$V = V_0 (\lambda/R)^{1/2} \dots\dots (1)$$

where R is the pathlength travelled, and corresponds to the expected power decay from a line source, $P = P_0 \lambda/R$, derived on the assumption that the propagation is unperturbed by the surface. The situation is sketched below:-



3. Insertion Loss of S.S.B.W. devices.

As usual, various theoretical approaches can be adopted to calculate the performance, including insertion loss, of S.S.B.W. devices. The most fundamental of these is to solve the complicated equations of motion in the (piezoelectric) medium subject to the appropriate elastic and electrical boundary conditions⁽⁶⁾. Milsom has recently applied this approach

to one of our S.S.B.W. devices, and his calculations are in respectable agreement with our measurements as regards centre frequency, bandshape and minimum insertion loss ($12\frac{1}{2}$ dB as compared with the measured value of 13 dB⁽⁸⁾). Unfortunately the detailed results were not available in time to be included in this paper, but will be published elsewhere at a later date.

Although this rigorous approach is valuable it gives little physical insight into the operation of S.S.B.W. devices, and leaves one guessing as to how, for example, the output I.D.T. is able to extract as much power as it does from a bulk wave which may have penetrated many wavelengths into the substrate. Fortunately there is an alternative approach to the problem which relies on general properties of waves, and which gives a great deal of insight into the operation of S.S.B.W. devices. This approach is based on antenna theory, and formed the basis of our initial qualitative understanding of the operation and insertion loss of S.S.B.W. devices, and which allowed us to design low-loss devices before the results of the rigorous approach were available⁽⁸⁾. Some of the elements of antenna theory, and its applications to S.S.B.W. devices are included in Appendix II.

By combining the rigorous approach and antenna theory one can develop a sound understanding of the operation of S.S.B.W. devices.

4. Frequency response of S.S.B.W. filters.

One of the attractions of SAW filters is the ease with which a bandpass filter with a required response may be designed by means of a simple Fourier transform. As an alternative to the use of Fourier transforms one may calculate the response of a given I.D.T. at any frequency by means of a phasor diagram (analogous to the computation of the response of an optical diffraction grating). Exactly the same procedure may be used to calculate the far-field intensity of a S.S.B.W. launched along the surface towards the output I.D.T. (This procedure is related to the calculation of the aerial gain in appendix II). Consequently we expect a 1:1 relationship between the response of a SAW filter and a S.S.B.W. filter operating under far-field conditions, and we have seen an example of this in Fig. 2(c). Similar results have been obtained on a SAW/S.S.B.W. T.V. filter with a more complicated pass-band. We should, however, point out two minor differences between SAW and S.S.B.W. filters:-

(i) In real devices it may well not be possible to operate under far-field conditions, in which case the R^{-2} factor in Eq (1) must be taken into account in the design.

(ii) In the case of SAW filters the "response" loosely referred to above actually relates to the series radiation resistance, R_s . In S.S.B.W. devices the I.D.T. launches bulk waves into the medium of the substrate at all frequencies above the centre of the S.S.B.W. response, and this gives rise to a "tail" on the high-frequency side of the R_s plot. Consequently there is no longer a 1:1 relationship between the filter response and the R_s plot, and this complicates the design of S.S.B.W. filters, especially when the I.D.T.'s are matched to the external circuit.

In general terms, however, we may safely conclude that the design of S.S.B.W. filters is only a little more complicated than the design of SAW filters.

5. Substrate orientations for S.S.B.W.

In seeking suitable orientations to support S.S.B.W. we naturally start by considering the most widely used SAW piezoelectric crystals, quartz, LiNbO_3 and LiTaO_3 . Following the discussion in Section 2, we shall concentrate on substrate orientations which support a horizontally-polarized shear wave. In addition, we would like the chosen orientations to satisfy as many as possible of the following criteria:-

- (a) Large piezoelectric coupling to the horizontally polarized shear wave (S.S.B.W.)
- (b) Zero beam steering both on the surface and into the volume (S.S.B.W.)
- (c) Zero temperature coefficient of delay (S.S.B.W.)
- (d) Zero piezoelectric coupling to SAW.
- (e) Zero piezoelectric coupling to other bulk waves, or negligible propagation of these waves for the reasons discussed in Section 2 and appendix I.

A family of cuts of quartz satisfying most of these criteria is the rotated y-cuts with propagation perpendicular to the x-axis. Here, the SAW coupling vanishes identically, all such cuts support an x-polarised shear wave which suffers no beam steering on the surface, and, for two small ranges of angle of the rotated y-cut, the temperature coefficient of delay is small. These ranges are $+30^\circ$ to $+40^\circ$ and -48° to -55° rotated y-cuts. These latter cuts are closely related to the well known AT and BT-cuts of conventional bulk-wave-oscillator crystal, but their roles are inverted because, in our new devices, the propagation (k -vector) is along the surface, rather than perpendicular to the surface.

By good fortune, it also happens that these cuts suffer a negligible beam steering into the volume of the material, and that the longitudinal wave response is not detectable, so that very 'clean' responses are obtained. Most of our recent studies, e.g. Fig. 2, have therefore concentrated on these two cuts of quartz, especially the AT cut ($+35.3^\circ$ rotated y-cut) which has essentially zero temperature coefficient of delay at room temperature in our devices, as discussed below.

Some experiments involving ST-quartz with propagation perpendicular to the x-axis have recently been reported by Yen et al.⁽¹³⁾. Since the ST-cut is the $+42^\circ$ rotated Y-cut, it is evident that this cut is one particular example of the family of cuts described above.

In our own investigations, we have measured the temperature coefficient of frequency (or delay) on the family of rotated Y-cuts (all with k perpendicular to the x-axis) covering the range $+33^\circ$ to $+38^\circ$, as these have inversion temperatures in the range of interest, see Fig. 4. We have also measured the effective velocity (deduced from the centre frequency of our S.S.B.W. response) and plotted the results in Fig. 5. Both these curves are in close agreement with the corresponding curves for bulk wave quartz crystal oscillator plates⁽¹⁴⁾, and serve to verify our interpretation of the principles of the S.S.B.W. devices.

It is important to add that the measurements in Figs 4 and 5 employed an effective transducer length of $N\lambda = 2500\lambda$, to ensure the excitation of as narrow an acoustic beam as possible. The angular spread of k -vectors (see sketch in section 2) is of order

$\Theta \approx (2/N)^{1/2}$ radians, i.e. $\Theta < 2^\circ$ in our device. In shorter transducers the results of Figs. 4 and 5 may be slightly modified because of the excitation of a wider range of k -vectors. The cuts shown in Figs 4(a) and 5 correspond to the BT-class of conventional bulk wave oscillator crystal.

The other range of orientations of interest (because of their small temperature coefficient of delay) is the -48° to -55° rotated y -cuts (again with k perpendicular to the x -axis). Preliminary measurements on two such cuts have yielded a superior temperature coefficient to the curves shown in Fig 4(a). This is expected, as the -48° to -55° cuts correspond to the AT-class of conventional bulk wave quartz crystal oscillator. These measurements are as yet incomplete, but those available are shown in Fig 4(b).

When we consider LiNbO_3 and LiTaO_3 , we find that although these materials have the same elastic symmetry as quartz they have very different piezoelectric symmetry. This is unfortunate as the family of rotated y -cuts with k perpendicular to the x -axis is no longer ideally suited to S.S.B.W. applications. We have therefore searched for other orientations in these materials, which satisfy most of the earlier criteria "by accident", rather than by symmetry as was the case for the rotated y -cuts of quartz. To date we have located two suitable cuts of LiTaO_3 . These employ the two pure shear bulk waves which propagate on the x -axis of LiTaO_3 in the infinite medium. One is the $+36^\circ$ rotated y -cut with velocity ~ 4210 m/s and k parallel to the x -axis. The other is the -54° rotated y -cut with velocity ~ 3460 m/s and k parallel to the x -axis. Neither has a particularly good temperature coefficient, but the former has a very high piezoelectric coupling constant for S.S.B.W. In each case the piezoelectric coupling to SAW is negligible⁽¹⁵⁾.

6. Applications of S.S.B.W. devices.

(i) Oscillators

The properties of S.S.B.W. devices discussed in the previous sections suggest that they are ideally suitable for use as the feedback element of a high frequency oscillator. In this role they extend the frequency coverage of SAW oscillators⁽⁹⁾ to about 1.6 times the previous upper limit, i.e. from about 1.5 GHz to 2.4 GHz. Fig 6 shows the response of two such devices on AT-quartz (with k perpendicular to the x -axis) and operating at 1.766 GHz and 2.323 GHz respectively. The principal experimental difficulty at these higher frequencies concerns direct e.m. breakthrough, and special packages are necessary to achieve satisfactory results.

The insensitivity to surface contamination, e.g. Fig.1, led us to inquire whether the new devices are less prone to long-term aging than are SAW devices. Experiments on two such oscillators have, indeed, shown a reduced aging (by a factor of 2-3) over the first 3 months. Such results are obviously preliminary and the interpretation is tentative, but they are, nevertheless, encouraging.

(ii) S.S.B.W. resonators

We have seen in Section 3 that the output I.D.T. is able to couple strongly to the S.S.B.W. despite the deep penetration of the wave into the substrate. In a similar manner an incident S.S.B.W. can be reflected efficiently from a periodic perturbation of the surface, e.g. by the I.D.T. or an array of grooves, such as is used in SAW resonators. We have therefore fabricated a S.S.B.W. device from one of Marshall's

"all-aluminium" SAW resonator patterns⁽¹⁶⁾. The resulting performance is shown in Fig. 7 and verifies that resonator-action is taking place. The Q of this device was a modest 1000, and obviously more work is necessary to provide devices with the inherent Q of the medium (which, incidentally, is higher than the inherent Q of a SAW resonator because of the lower propagation losses of bulk waves).

(iii) Harmonic operation of S.S.B.W. devices.

It is well-known that the I.D.T. or modifications thereof can be used to excite SAW at the fundamental and various harmonic frequencies⁽¹⁷⁾. For the simplest I.D.T., the first strong harmonic is the 5th harmonic. We have found the same result to be true for S.S.B.W., and this leads to the possibility of fabricating devices at even higher frequencies than described in (i) above.

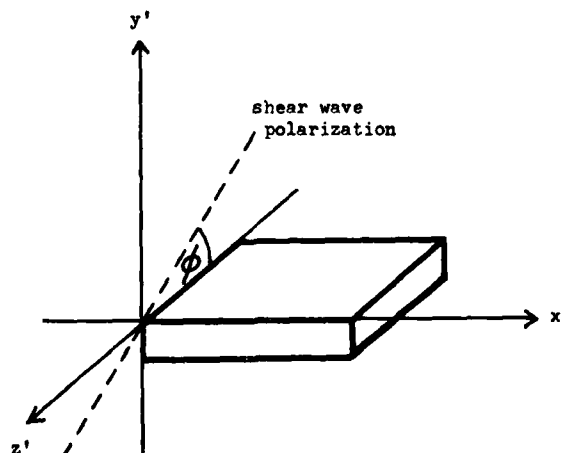
7. Conclusions

We have described in some detail the properties of a new class of planar bulk acoustic wave delay line which has many properties in common with SAW devices, but certain advantages over SAW for particular applications. These applications include high frequency oscillators and filters, and we have demonstrated the production of such devices at fundamental frequencies up to 2.3 GHz. Although fundamental studies of these devices are as yet incomplete, the results obtained to date are most promising.

In the immediate future we aim to extend these studies to cover the temperature coefficients of the full range of S.S.B.W. devices which are analogous to the conventional AT-cut of bulk wave oscillator. We shall also extend our studies on LiTaO_3 to see if it is possible to exploit the higher piezoelectric coupling of S.S.B.W. on this material in filter applications. It is also clear from the text that there are still many aspects of S.S.B.W. which warrant attention and we hope to report on some of these at a later date.

Appendix I

We are concerned with the surface stresses arising from the propagation of bulk acoustic waves in a finite substrate in the form of a plate. Let the plate normal be parallel to the y' -axis, and let propagation be in the x' -direction:



Assume that the aperture of the acoustic wavefront is many wavelengths wide (i.e. in the z' -direction) so that end effects may be neglected. The surface boundary conditions that must be satisfied if the wave is to propagate freely are that the stresses over the y' -plane should vanish, $T_{1'2'} = T_{6'} = 0$; $T_{2'2'} = T_{2'} = 0$ and $T_{3'2'} = T_{4'} = 0$.

(a) Isotropic materials

Consider first a longitudinal wave. This has only one strain component $S_{1'1'} = S_1$, and this is coupled to $T_{2'}$, $T_{4'}$, $T_{6'}$ through elastic constants $c_{1'2'}$, $c_{1'4'}$, and $c_{1'6'}$, respectively. For isotropic materials $c_{1'4'} = c_{1'6'} = 0$, but $c_{1'2'} = c_{12} \neq 0$ ⁽¹⁸⁾. Therefore the longitudinal wave will not, in general, satisfy the surface boundary conditions.

Consider next a shear wave polarized at an angle ϕ to the z' -axis as indicated in the sketch. Such a wave has strain components $S_{1'3'} = S_5 = \frac{1}{2} \cos \phi$ and $S_{1'2'} = S_6 = \frac{1}{2} \sin \phi$. These components are coupled to $T_{2'}$, $T_{4'}$, and $T_{6'}$ via the following elastic constants:-

Strain/Stress	$T_{2'}$	$T_{4'}$	$T_{6'}$
S_5	$c_{2'5'}$	$c_{4'5'}$	$c_{5'6'}$
S_6	$c_{2'6'}$	$c_{4'6'}$	$c_{6'6'}$

For an isotropic material all these elastic constants vanish except $c_{6'6'} = c_{44} \neq 0$. It is therefore clear that the boundary conditions are only satisfied if $S_6 = 0$, i.e. if $\phi = 0$ and the shear wave is horizontally-polarized, in our previous notation.

(b) Rotated y-cut plates of materials in Laue group RI

This group includes quartz, LiNbO_3 and LiTaO_3 . Let the angle of rotated y-cut be θ . The elastic constants in the new coordinate system have been given by Sykes⁽¹⁹⁾. As in the case of isotropic materials above consider a shear wave polarized at an angle ϕ to the z' -axis and propagating in the x -direction. The same coupling constants appear as in the table above, but now $c_{6'6'} \neq 0$ and $c_{5'6'} = 0$ only for 2 particular values of θ . The condition $c_{6'6'} = 0$ requires that the shear wave be horizontally polarized. Examination of the determinantal equation of wave propagation in the rotated coordinate system shows that the determinant factorizes for precisely those values of θ that make $c_{5'6'} = 0$. In these circumstances the shear waves are polarized exactly along the y' - and z' - axes, showing that there are in all materials two particular values of θ which support horizontally-polarized shear waves which exactly satisfy the surface-boundary value conditions. In quartz the appropriate cuts are the AT (+11°) and BC (-59°) cuts, which are of course close to the XT and AT cuts, and BT cuts respectively.

The corresponding cuts in LiTaO_3 are the +36° and -54° rotated y-cuts as discussed in section 5.

Let us now consider propagation of an x -polarized shear wave in the z' -direction. (It can be shown that such a wave always exists in the infinite medium). The only strain component involved is S_5 , and the boundary value conditions are only strictly satisfied if $c_{5'6'} = 0$, i.e. for precisely the same rotated y-cuts as found above for x -propagation. It is evident from our experiments (Figs 2, 4, 5) that for the ranges of rotated y-cut studied (+35° to +38°) and (-48° to -55°) these are sufficiently close to the AC and BC cuts to allow propagation over significant distances. This is consistent with the fact that $c_{5'6'}$ remains small for all angles of rotated y-cut quartz.

Appendix II

In this appendix we study the application of antenna theory⁽²⁰⁾ to the excitation and reception of S.S.B.W. by an I.D.T. We know that the boundary has little or no effect on the propagation of those waves of interest, viz horizontally polarized S.S.B.W., and therefore we may regard the I.D.T. as an acoustic aerial operating in a half-space. (It will make little difference to the argument if we imagine that another half-space is bonded to the substrate, and that the I.D.T. is therefore buried in the medium). By analogy with e.m. aerial theory, consider two ideal ISOTROPIC acoustic aerials, I, separated by a large distance, R , in an isotropic medium that so each is in the far field of the other. Imagine that each aerial has its acoustic radiation resistance tuned and matched to the source/load impedance. Then, for an input power P_{IN} , the power density falling on the output aerial is $P_{IN}/4\pi R^2$. It is a basic result of aerial theory⁽²⁰⁾ that the effective cross-sectional area of the receiving isotropic aerial is $\lambda^2/4\pi$. Therefore the output power is given by

$$P_{OUT} = P_{IN} \times \frac{\lambda^2}{(4\pi R)^2} \dots\dots\dots(2)$$

In the case of an I.D.T. of length $N\lambda$, it behaves "on resonance" like a bidirectional endfire array aerial which radiates a main lobe of semiangle $\theta \approx (2/N)^{1/2}$ and with aerial gain in the forward direction, $G \sim N/2$. In these circumstances the expression for P_{OUT} in Eq (2) must be multiplied by $G_1 G_2$. If we extrapolate these far field results to the case of interest to SAW or S.S.B.W. oscillators, viz two identical I.D.T.'s of length $N\lambda$ and centre-to-centre separation $N\lambda$ ⁽⁹⁾, we obtain from Eq(2) an insertion loss of ~28dB. Actually this calculation overestimates the loss because in reality the aperture of the I.D.T.'s is so large that as far as this dimension is concerned, the aerials are in each other's near field. In practice we have made devices on quartz and LiTaO_3 with losses as low as 13 dB.

The use of this antenna theory approach not only allows us to design S.S.B.W. devices with the minimum attainable insertion loss, but it gives us a crude estimate of this loss, and of the distribution of acoustic energy within the device.

References

- (1) W. Mortley, U.K. Patent 988102 (1962).
- (2) R. M. White and F. W. Voltmer, Appl Phys Lett. 7, 314 (1965).
- (3) See for example, IEEE Special issue, Trans MTT-17 (1969).
- (4) D. P. Morgan (Ed). 'SAW passive interdigital devices' IEE reprint series 2 (1976).
- (5) R. F. Mitchell, Proceedings of IEEE Ultrasonics Symposium (IEEE cat 74 CHO 896-1 SU) 313 (1974).
- (6) R. F. Milsom, Wave Electronics 2 64 (1976).
- (7) M. F. Lewis, U.K. patent 1451326; U.S.A. patent 3950713.
- (8) T. I. Browning and M. F. Lewis, Electronics Letters 13 128 (1977).
- (9) M. F. Lewis, Ultrasonics, 12, 115 (1974).
- (10) R. S. Wagers, IEEE Trans SU-23, 113 (1976).
- (11) T. C. Lim and G. W. Farnell, J.A.P. 39, 4319, (1968).
- (12) H. Kolsky "Stress Waves in Solids". Dover publications (New York, 1963) c.p. 48.
- (13) K. H. Yen, K. L. Wang and R.S. Kagiwada, Electronics Letters 13 37 (1977).
- (14) W. P. Mason "Piezoelectric Crystals and their Application to Ultrasonics". D.Van Nostrand Co.(1950) p.99 especially figs 6.12 and 6.13
- (15) A. J. Slobodnik Jr., et al "Microwave Acoustics Handbook, Vol 1A" (1973) p.94.
- (16) F. G. Marshall, Proceedings of the IEEE Ultrasonics Symposium (IEEE Cat 75 CHO 994-4 SU) 290 (1975).
- (17) H. Engan, IEEE Trans ED-16, 1014 (1969).
- (18) K. Brugger, J.A.P. 36, 759 (1965).
- (19) R. A. Sykes. B.S.T.J. 23 (1944).
- (20) "Antennas: Theory and Practice", I.S. Sokolnikoff (Wiley, New York, 1952).

Acknowledgement

The authors wish to acknowledge useful discussions with various members of these laboratories, and the technical assistance of Geoff Fynn, Alex Young and Len Hunt in producing and measuring the properties of so many special devices.

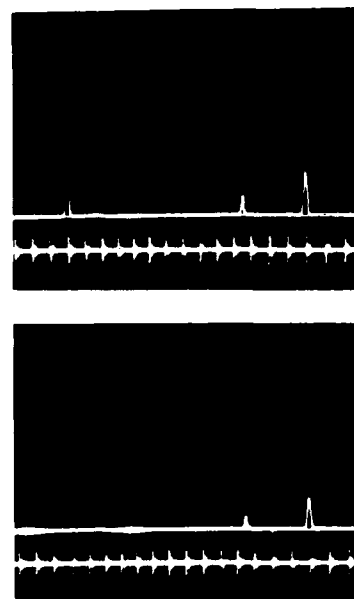


Fig.1. (a) Transmitted power versus frequency of a single-mode SAW oscillator delay line filter on AT-quartz with propagation along the x-axis. The SAW response is at 170 MHz, and the fast transverse and longitudinal bulk wave responses at 275 and 309 MHz respectively. Markers in the lower trace are separated by 10 MHz.

(b) Frequency response of the same device as (a) but with absorber added in the 15λ wide gap between transducers.

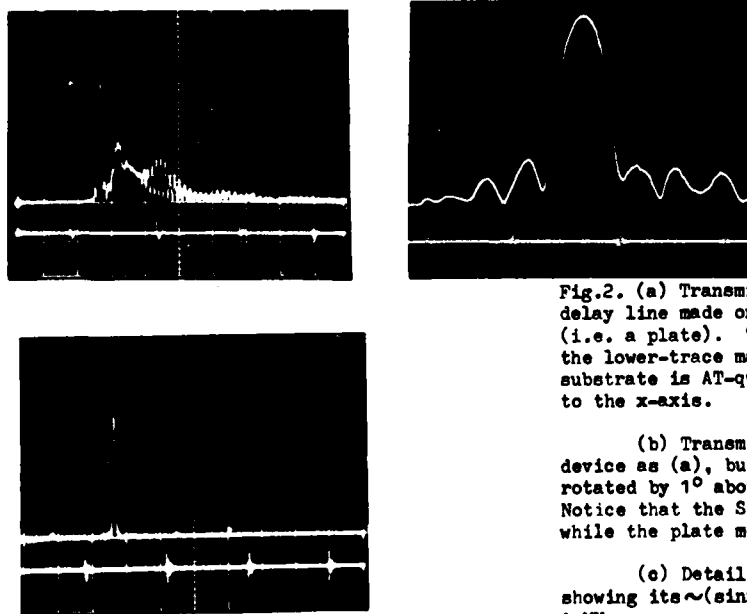


Fig.2. (a) Transmission characteristics of a S.S.P.W. delay line made on a substrate with parallel faces (i.e. a plate). The centre frequency is 154 MHz and the lower-trace markers are separated by 10 MHz. The substrate is AT-quartz with propagation perpendicular to the x-axis.

(b) Transmission characteristics of the same device as (a), but with the bottom face of the plate rotated by 1° about the direction of propagation. Notice that the S.S.B.W. response is unaffected, while the plate modes are eliminated from the response.

(c) Detailed photograph of S.S.B.W. response showing its $\sim(\sin x/x)$ form. Markers are separated by 1 MHz.

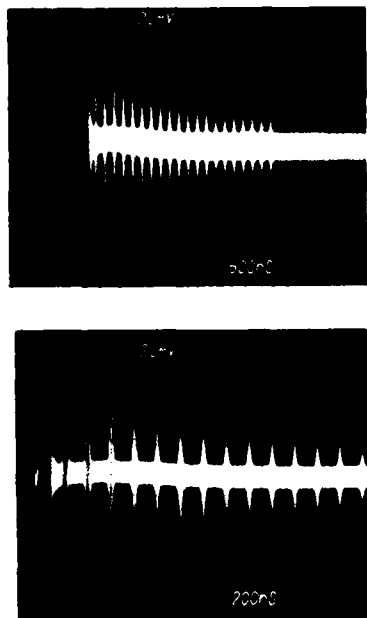


Fig. 3. (a) Measured impulse response of the device sketched in section 2. Notice that the voltage varies as indicated in Eq (1).

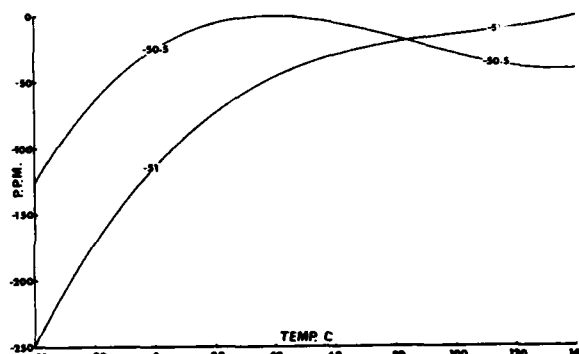
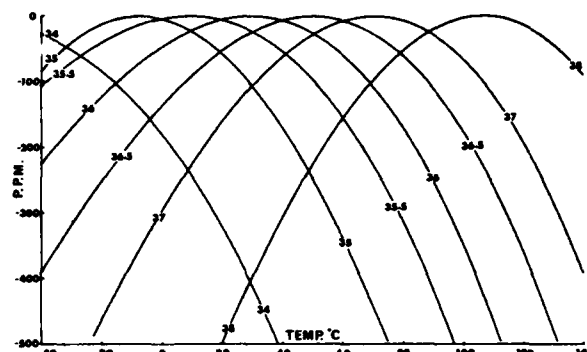


Fig.4. (a) Measured variation of the frequency with temperature of a series of S.S.B.W. devices fabricated on rotated y-cut quartz with propagation perpendicular to the x-axis. The angles of rotated y-cut are indicated on each plot. It is important to note that these measurements employed 2500 Å transducers, as discussed in the text.

(b) Preliminary measurements of the frequency versus temperature coefficients of two S.S.B.W. devices on rotated y-cuts which are analogues of the AT-type of bulk wave device. Propagation is perpendicular to the x-axis.

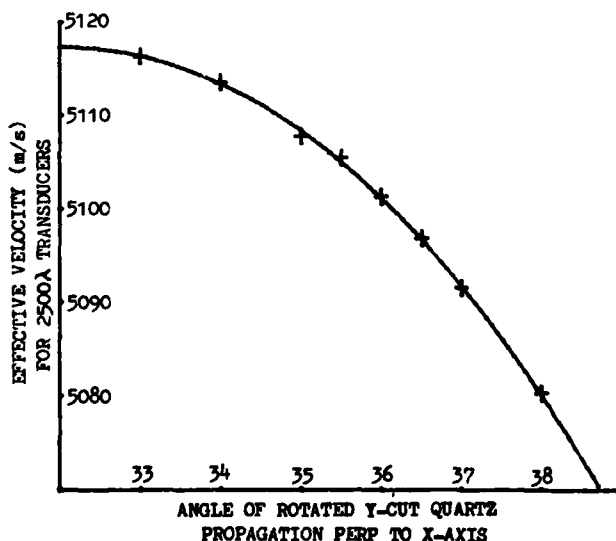


Fig.5. Effective velocity versus orientation for the S.S.B.W. devices whose temperature coefficients are shown in Fig.4.(a).

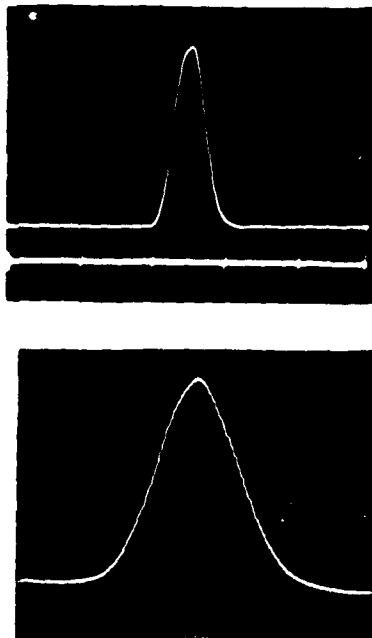


Fig.6. (a) Frequency response of a 1.766 GHz S.S.B.W. oscillator filter made on AT-quartz with propagation perpendicular to the x-axis. Total scan illustrated:- 50 MHz.

(b) Frequency response of a 2.323 GHz S.S.B.W. filter on the same substrate orientation as (a). Total scan illustrated:- 20 MHz.

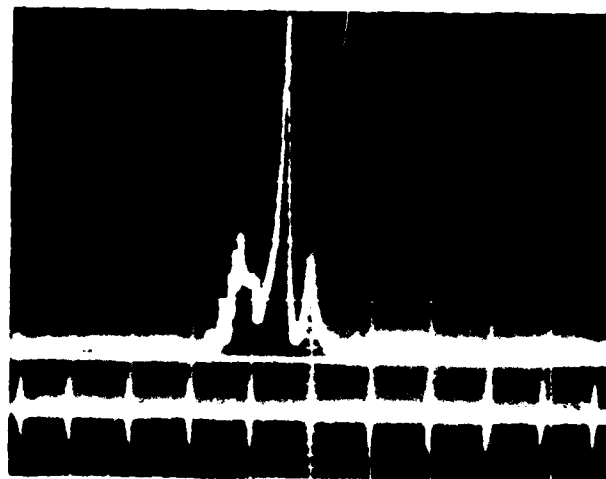


Fig.7. Response of an experimental S.S.B.W. two-port resonator. The substrate is AT-quartz with propagation perpendicular to the x-axis. The full width at -3dB points is 130 kHz. The corresponding width for the transducers alone is 1.5 MHz. The centre frequency is 151 MHz, and markers in the lower trace are separated by 1MHz.

INTERDIGITAL TRANSDUCERS--

A MEANS OF EFFICIENT BULK WAVE EXCITATION

K. H. Yen, K. L. Wang, R. S. Kagiwada, and K. F. Lau
TRW Defense and Space Systems Group
Redondo Beach, California .

Summary

Interdigital transducers (IDT) can efficiently generate and detect surface acoustic waves (SAW) as well as bulk waves. The spurious generation of bulk waves is undesirable for SAW device applications. Bulk-wave excitation in SAW devices has been theoretically and experimentally investigated, and some techniques for their reduction and suppression have been suggested. However, bulk waves can be used to play a useful role in device applications. We present the experimental results for the efficient excitation and detection of a bulk wave by SAW IDT's on rotated Y-cut quartz.

Both apodized and unapodized interdigital transducers were used for the delay lines. The transducers were operated at fundamental and higher harmonic frequencies. The experimental results show that the efficiency of excitation is comparable to that of SAW's on rotated Y-cut quartz. The responses were free of spurious responses. These efficient excitations of bulk waves do not require a piezoelectric overlay such as ZnO on the transducer to enhance the coupling efficiency. The center frequency is about 1.6 times higher than that of an X-propagated SAW device using an identical IDT.

It has been experimentally determined that these bulk waves propagate close to the surface of the substrate. The energy is beamed directly from one transducer to the other. These shallow bulk acoustic waves (SBAW) are insensitive to the surface contamination and crystal dimension. They respond at higher frequencies and will be more temperature-stable than SAW devices. Thus, devices using SBAW can be expected to combine some of the best features of both the bulk crystal technology and the surface acoustic wave technology.

Introduction

Interdigital transducers (IDT) can be used to generate and detect surface acoustic waves (SAW) and bulk waves.¹⁻⁷ For SAW devices, the generation of bulk waves is undesirable. This is because bulk waves give rise to spurious signals and unwanted loss. Many theoretical and experimental investigations have been done in this area in order to develop techniques to suppress them.^{5,6,7} For example, it has been found that bulk waves reflected from the bottom face can usually be removed by tapering the substrate. But bulk waves traveling at a shallow angle to the surface and beaming directly to the output transducer are difficult to deal with. However, we have recently realized that such bulk waves traveling at a shallow angle to the surface can also be utilized to play a useful role in device applications.⁸ By choosing the correct piezoelectric material and crystallographic orientation, we have produced a new family of planar bulk wave devices employing SAW interdigital transducers.

The successful operation of the planar bulk wave devices is due to the efficient generation and detection of the "Shallow Bulk Acoustic Waves" (SBAW) by

interdigital transducers. This type of bulk wave travels close to the surface of the substrate and is beamed directly from one transducer to the other. In the case of rotated Y-cut quartz, these SBAW's can be generated and detected at frequencies 1.6 times higher than the corresponding SAW's using identical IDT's. This greatly reduces the linewidth resolution for IDT's. Furthermore, experimental evidence indicates that SBAW can be manipulated in ways similar to SAW's. Thus, devices using SBAW can be expected to combine desirable features of both the bulk crystal technology and the surface acoustic wave technology.

The Nature of the Shallow Bulk Acoustic Waves

Recently, Milson, et al. have shown that horizontally polarized shear waves can be generated and detected on PZT-4 and 41°-rotated YX-LiNbO₃ by interdigital transducers.⁹ These waves have simple transverse particle motion and high coupling coefficient. They travel in the direction of Bleustein-Gulyaev mode orientation. On PZT-4, with a free surface between transducers, the bulk shear waves propagate with the main lobe of radiation almost parallel to the surface of the substrate. On 41°-rotated YX-LiNbO₃, the presence of the bulk shear waves causes passband distortion of Bleustein-Gulyaev wave devices. However, we have recently shown that these shallow bulk acoustic waves can also be utilized to play a useful role in device applications.⁸

In order to minimize spurious responses, the basic conditions for shallow bulk acoustic waves excitation are:

1. Zero piezoelectric coupling to surface acoustic waves.
2. Nonzero piezoelectric coupling to bulk acoustic waves.

For example, condition (1) is satisfied by the family of Y-cut quartz with the SBAW traveling 90° from the X-axis.¹⁰ But there is no SBAW excitation in this direction because the second condition is not met.¹¹ On the other hand, SBAW's can be generated by SAW IDT's on rotated Y-cut quartz in which both conditions are satisfied.

The surface amplitude of the shallow bulk acoustic waves is a Hankel function which decays as the inverse square root of distance.⁹ The output power therefore depends on transducer separation L and varies as $P = P_0 \lambda / L$. Hence, the insertion loss of SBAW devices is a function of transducer separation.

The shallow bulk acoustic wave is a pure shear wave. Thus, all the shear type crystal orientations may be used to generate these waves. For example, both AT- and BT-cut quartz can be used for SBAW devices. These devices will have better temperature stability than do SAW's.

Advantages of Shallow Bulk Acoustic Wave Devices

SBAW devices combine many good features of both the bulk acoustic wave device and surface acoustic

wave device. More specifically, the advantages of the shallow bulk acoustic wave devices over normal bulk wave devices are:

- Higher frequency of operation
- Weighted frequency response (time domain)
- Operating frequency independent of crystal dimensions
- Free of surface contaminations
- Lower spurious response
- Planar structure
 - mechanically rugged structure
 - good mechanical and thermal contact
 - readily fabricated using semiconductor technique (made easily and cheaply)
 - readily integrated with microelectronic components

Advantages of shallow bulk acoustic wave devices over surface acoustic wave devices are:

- Higher operating frequency (orientation-dependent)
- High temperature stability (orientation-dependent)
- Lower spurious response
- Free of surface contaminations

A temperature-stable orientation is already available in AT-cut quartz. The AT-cut bulk crystal cut has a temperature coefficient superior to that of ST-cut quartz, which is the best temperature-stable material for SAW devices.

The outlook is favorable for finding materials which possess orientations that are highly temperature-stable and have high conversion efficiency.

Experimental Arrangement and Results-- SBAW Delay Lines

Until recently, bulk wave excitations by interdigital transducers have been studied only with the aim of trying to eliminate them. We have recently pointed out that bulk waves can be utilized to construct a wide variety of devices.⁸ These shallow bulk acoustic wave devices will have device performances which will exceed those of bulk crystals and surface acoustic wave devices.

The experiments on shallow bulk acoustic wave devices were done on ST-cut quartz and 34°36' rotated Y-cut quartz. These quartz substrates were polished on one surface with the standard Valtec surface acoustic wave finish. All the quartz samples were X-rayed to verify the crystallographic orientation. The same photomasks for the delay lines were used to fabricate in both the X-direction and the 90° rotated direction on the same quartz substrate. The X-propagation direction excited the surface acoustic waves, while the 90° rotated direction excited the shallow bulk acoustic waves. In these experiments, existing SAW delay line photomasks were used in fabricating these devices. No attempt was made to optimize the shallow bulk acoustic wave's response. A more thorough theoretical and experimental analysis is presently being undertaken.

The first delay line used to excite shallow bulk acoustic waves consisted of two conventional SAW interdigital transducers. One transducer has a total of 74 finger pairs, the other has 106 finger pairs. Each transducer has a periodicity of 12,628 μm , and the center-to-center separation between transducers is 1.34 mm. The delay line was fabricated 90° from the X-axis. The response of this device is shown in Figure 1. This delay line was designed to operate at 250 MHz on X-propagated ST-cut quartz. When the propagation direction was rotated 90° with respect to the X-axis, the delay line responded at 393.65 MHz (~ 1.6 times higher). The unmatched insertion loss is about 16 dB (same as the propagated SAW delay line). The frequency response shown in Figure 1 is not a $(\sin X/X)^2$ function. The asymmetry in frequency response is due to the fact that the plate modes exist in the substrate. These plate modes can be eliminated by tapering or sandblasting the bottom surface of the substrate.

The shallow bulk acoustic waves have no spurious responses. This is demonstrated in Figure 2, which shows the response of the same device with a frequency from 200 to 600 MHz. The measured temperature coefficient of delay of this SBAW delay line is about 28 ppm/°C. This is already smaller than many commonly used SAW substrates such as YZ-LiNbO₃ and YZ-LiTaO₃. On rotated 34°36' Y-cut quartz, the SBAW excited by the same delay line has the frequency response shown in Figure 3. The insertion loss is about 14 dB (same as X-propagated SAW). The center frequency is again about 1.6 times higher than that of X-propagated SAW.

Figure 4 shows the frequency response of a SBAW delay line using two 45 finger-pair transducers. The separation between transducers was 0.7 mm. The insertion loss is about 14 dB, which is about 14 dB less than that of X-propagated SAW's. The lower insertion loss is due to the fact that SBAW has a higher piezoelectric coupling constant.

Thinned electrode transducers with and without apodization have also been utilized to excite shallow bulk acoustic waves.¹² A delay line consisting of two thinned electrode transducers, one without apodization and the other with apodization, was fabricated on ST-cut quartz with the propagation direction 90° from the X-axis. One transducer has 37 sections and 5 fingers per section. The other transducer has 40 sections and 8 fingers per section. The latter transducer was also weighted with a Hamming function. The periodicity of the electrodes was 2.84 μm and the center-to-center separation between transducers was 6.25 mm. This delay line was designed to operate at 277.5 MHz on X-propagated ST-cut quartz. The frequency response of the 90° rotated transducer is shown in Figure 5. The center frequency is 439 MHz and the insertion loss about 29 dB. The frequency response for the SAW delay line propagated along the X-direction was in good agreement with theory. Its untuned insertion loss was 18 dB.

The frequency response of other Hamming function weighted delay lines is shown in Figure 6. The delay line consists of Hamming function weighted 4-finger per period transducers of 160 finger pairs and 3-finger per period transducers of 80 pairs.¹³ The center-to-center separation is 7.082 mm. The untuned insertion loss is about 30 dB, which is about 6 dB higher than that of X-propagated SAW. The spurious response shown in Figure 6 is due to plate modes, which can be eliminated by tapering the bottom face of the substrate.

Multi-electrode transducers operating on their higher harmonics were also successfully used to

generate and detect shallow bulk acoustic waves. A delay line consisting of a 4-finger transducer and a 3-finger transducer was fabricated on ST-cut quartz. The 4-finger transducer had 37 periods and was excited at its 9th harmonic. The 3-finger transducer had 63 periods and operated at its 8th harmonic.¹³ The center-to-center separation between transducers was 3.16 mm. A surface acoustic wave delay line fabricated along the X-direction responded at 500 MHz. Its measured insertion loss was 40 dB, which agreed very well with calculated figures. The response of the shallow bulk acoustic wave, rotated 90° from the X-direction, is shown in Figure 7. The center frequency was about 790 MHz and the untuned insertion loss was 42 dB.

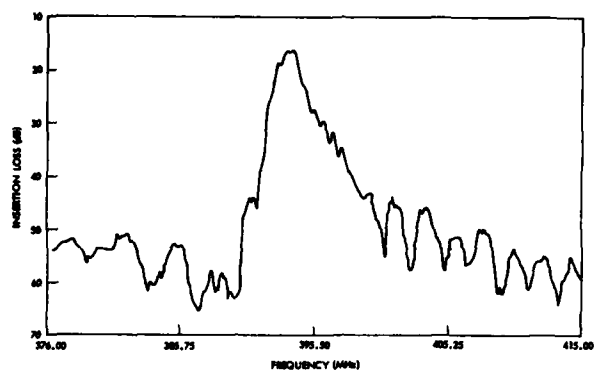
Finally, Figure 8 shows the frequency response of a 1.77 GHz SBAW delay line. This delay line consisted of two thinned electrode transducers. One has 10 sections, 4 finger pairs per section; the other has 14 sections, 4 finger pairs per section. The periodicity of the electrodes was 0.7 μ m and the center-to-center separation between transducers was .5 mm. The delay line was operated at fundamental frequency. The insertion loss is about 4 dB less than that of X-propagated SAW which has a center frequency of 1.1 GHz. This delay line clearly shows the advantage of using SBAW at high frequencies. It greatly reduces the finger width resolution of the interdigital transducers.

Conclusion

The efficient excitation and detection of shallow bulk acoustic waves described here does not require a piezoelectric overlay film such as ZnO on the transducers to enhance the coupling efficiency. The center frequency of operation on rotated Y-cut quartz is about 1.6 times higher than that of SAW's using identical interdigital transducers. There are many advantages of SBAW devices over both bulk waves and SAW devices. One is very optimistic about finding materials which possess orientations that are highly temperature-stable and have high conversion efficiency. SBAW oscillators using AT-cut quartz should have device performance better than that of SAW oscillators using ST-cut quartz. Because SBAW travels close to the surface of the substrate, one would expect that many SAW device applications can be carried over to SBAW's.

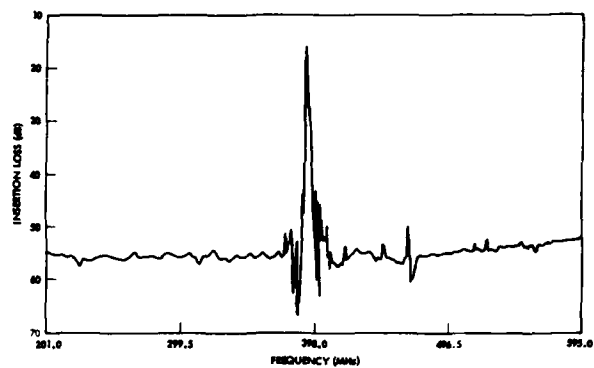
References

1. R. S. Wagers, "Analysis of Acoustic Bulk Mode Excitation by Interdigital Transducers," *Appl. Phys. Letters*, Vol. 24, No. 9, pp. 401-403, 1 May 1974.
2. R. S. Wagers, "Plate Mode Coupling in Acoustic Surface Wave Devices," *IEEE Sonics and Ultrasonics*, Vol. SU-23, No. 2, pp. 113-127.
3. R. Milsom, J. Heighway, N. Reilly, and M. Redwood, "Comparison of Exact Theoretical Predictions and Experimental Results for Interdigital Transducers," *Proc. of IEEE Ultrasonic Symposium*, Milwaukee, Wisconsin, 1974, pp. 406-411.
4. M. R. Daniel, "Acoustic Radiation from a High Coupling Cut of Lithium Niobate," *J. of Appl. Physics*, Vol. 44, No. 7, pp. 2942-2945, July 1973.
5. M. R. Daniel, P. R. Emtage, and T. deKlerk, "Acoustic Radiation by Interdigitated Grids on LiNbO₃," *Proc. of IEEE Ultrasonic Symposium*, Boston, Massachusetts, 1972, pp. 392-395.
6. R. F. Mitchell, "Spurious Bulk Wave Signals in Acoustic Surface Wave Devices," *Proc. of IEEE Ultrasonic Symposium*, Los Angeles, California, 1975, pp. 469-470.
7. A. L. Nalamwar, M. Epstein, and R. S. Wagers, "Efficient Excitation of Shear Bulk Waves by Thin-Film ZnO ID Transducers," *Proc. of IEEE Ultrasonic Symposium*, Los Angeles, California, 1975, pp. 469-470.
8. K. H. Yen, K. L. Wang, and R. S. Kagiwada, "Efficient Bulk Wave Excitation on ST-Quartz," *Electronics Letters*, Vol. 13, pp. 37-38, 1977.
9. R. F. Milson, N. H. C. Reilly, and M. Redwood, "Analysis of Generation and Detection of Surface and Bulk Acoustic Waves by Interdigital Transducers," *IEEE Trans. on Sonics and Ultrasonics*, SU-24, No. 3, pp. 147-166, May 1977.
10. A. J. Slobodnik, Jr., E. D. Conway and R. T. Delmonico, "Microwave Acoustic Handbook," Vol. 1A, *Surface Wave Velocities*, AFCL-TR-73-0597.
11. B. A. Auld, "Acoustic Fields and Waves in Solids," Vol. 1, p. 308.
12. K. R. Laker, A. J. Budreau and P. H. Carr, "Interconnecting SAW Filters for Low Loss Frequency Multiplexers and Frequency Synthesizers," *Proc. of IEEE Ultrasonic Symposium*, Milwaukee, Wisconsin, 1974, pp. 161-163.
13. S. J. Kerbel, "Design of Harmonic Surface Acoustic Wave (SAW) Oscillators Without External Filtering and New Data on the Temperature Coefficient of Quartz," *Proc. of IEEE Ultrasonic Symposium*, Milwaukee, Wisconsin, 1974, pp. 276-281.



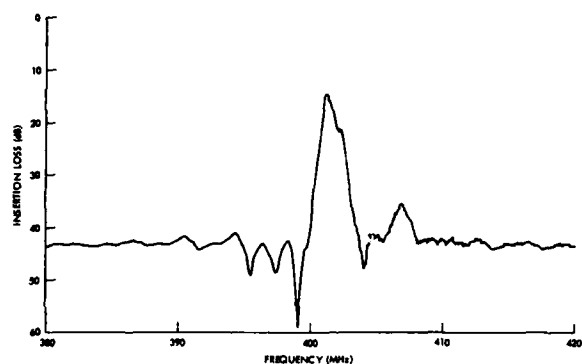
FREQUENCY RESPONSE OF THE SAW DELAY LINE USING CONVENTIONAL TRANSDUCERS ON ST-QUARTZ

FIGURE 1.



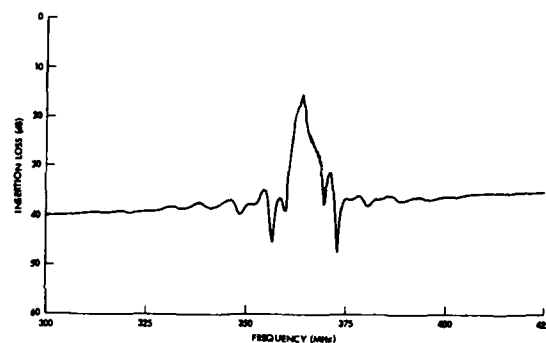
FREQUENCY RESPONSE OF THE SAW DELAY LINE USING CONVENTIONAL TRANSDUCERS ON ST-QUARTZ

FIGURE 2.



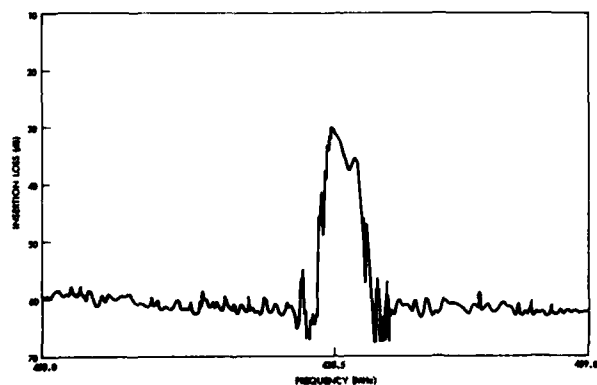
FREQUENCY RESPONSE OF THE SAW DELAY LINE USING CONVENTIONAL TRANSDUCERS ON 34° 36' ROTATED Y-CUT QUARTZ

FIGURE 3.



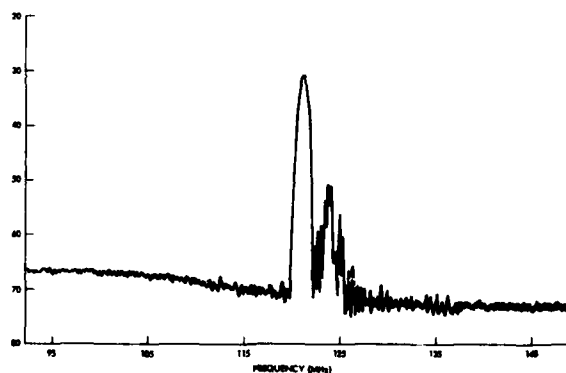
FREQUENCY RESPONSE OF THE SHALLOW BULK ACOUSTIC WAVE DELAY LINE USING 40 PAIRS OF INTERDIGITAL TRANSDUCERS ON ST-QUARTZ

FIGURE 4.



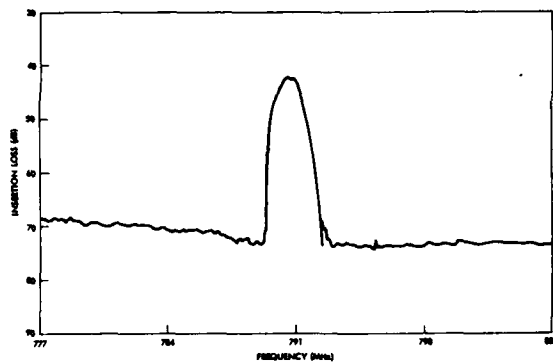
FREQUENCY RESPONSE OF THE SAW DELAY LINE USING HOPPING FUNCTION RELOADED THIN-FILM ELECTRODE TRANSDUCER ON ST-QUARTZ

FIGURE 5.



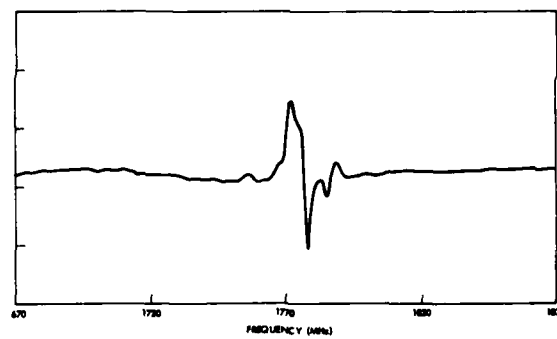
FREQUENCY RESPONSE OF THE SAW DELAY LINE USING HOPPING FUNCTION RELOADED TRANSDUCERS ON ST-QUARTZ

FIGURE 6.



FREQUENCY RESPONSE OF THE SAW DELAY LINE USING MULTIELECTRODE TRANSDUCERS ON ST-QUARTZ

FIGURE 7.



FREQUENCY RESPONSE OF THE 1.77 MHz SAW DELAY LINE ON ST-QUARTZ

FIGURE 8.

SURFACE ACOUSTIC WAVE SCATTERING FROM A GROOVE IN Y-Z LiNbO₃

S.D. Wu and H.S. Tuan
Department of Electrical Engineering
State University of New York at Stony Brook

Summary

The previously reported treatment of SAW scattering from a groove employs a substrate which is purely elastic and isotropic. Using the theory to study a groove in a Y-Z LiNbO₃, it is necessary to replace the piezoelectric substrate by an equivalent isotropic one with some equivalent material parameters. Although this approach simplifies the analysis of a complicated problem, however, it does not put the scattering characteristics "directly" in terms of the basic material constants of the crystal substrate. The present paper is directed to overcome the aforementioned shortcoming. It presents a theory of SAW scattering from a shallow groove in a Y-Z LiNbO₃, which includes both the elastic anisotropy and piezoelectric effects.

The guided surface mode in a LiNbO₃ is called a generalized Rayleigh wave due to the transverse decaying constants being complex. In this paper, an incident generalized Rayleigh wave is assumed to be propagating along the Z axis of a Y cut LiNbO₃, and incident on a shallow groove. The SAW scattering properties are studied by using a perturbation approach. The total acoustic and electric fields may be represented by an infinite sum of different order of scattered fields and the incident SAW. Boundary conditions needed for each order of the scattered fields are obtained systematically from the perturbation calculation. These conditions correspond to induced equivalent elastic and electric sources on an ungrooved surface. Thus, the determination of each order scattered fields is transformed into a problem of acoustic wave excitation with prescribed source distributions, which may be solved by the usual Fourier transform technique.

In this paper, the general formulation of the SAW scattering from a groove in a Y-Z LiNbO₃ is presented. The first order scattered fields are studied. Numerical evaluation of the reflection coefficient for a rectangular-like groove is obtained which agrees quite closely with experimental observations.

Introduction

Based on the scattering properties of a single groove, an array of shallow grooves can be used as a coherent surface acoustic wave (SAW) reflector. Using this property, an important class of SAW devices have been developed recently. A few of the well-known examples are the pulse compression filters, bandpass filters and high Q resonators. Since a single groove is a building block for an array, a thorough understanding of its characteristics is both important and useful. The previously reported treatment [1-3] of SAW scattering from a groove employs a substrate which is purely elastic and isotropic. Using the theory to study a groove in a Y-Z LiNbO₃, it is necessary to replace the piezoelectric substrate by an equivalent isotropic one with some equivalent material parameters [4]. Although this approach simplifies the analysis of a very complicated problem, however, it does not put the scattering characteristics "directly" in terms of the basic material constants of the crystal substrate. The pre-

sent paper is directed to overcome the aforementioned shortcoming. It presents a theory of SAW scattering from a shallow groove in a Y-Z LiNbO₃, which includes both the elastic anisotropy and piezoelectric effects.

The propagation characteristics of SAW in a piezoelectric substrate was first treated by Coquin and Tiersten [5] in connection to the problem of SAW excitation in Quartz, then by Tseng and White [6] on the basal plane of hexagonal crystals and by Campbell and Jones [7] in the general direction of a trigonal crystal such as LiNbO₃. The guided surface wave is called a generalized Rayleigh wave due to the transverse decaying constants being complex for both elastic displacements and electric field components. The scope of the present paper is confined to a formulation of the scattering problem from a single groove in a piezoelectric substrate with an incident generalized Rayleigh wave. Numerical computation of the reflection coefficient is also carried out which agrees well with experimental observations [8].

The Incident Rayleigh Wave

The geometry of the problem is shown in figure 1. A generalized Rayleigh wave is incident from left to right along the Z axis on the surface of a Y-cut LiNbO₃ substrate which occupies the space y > 0. The coordinate system x, y, z is chosen to coincide with the crystal axes X, Y and Z such that the stiffness constants, piezoelectric stress constants and permittivity constants of the crystal remain unchanged. A two dimensional problem is considered so that there is no variation in the x direction. The linear piezoelectric constitutive relation and the equation of motion for LiNbO₃ are

$$T_{ij} = C_{ijkl} S_{kl} - e_{ijk} E_k \quad (1)$$

$$D_i = e_{ikl} S_{kl} + \epsilon_{ik} E_k \quad (2)$$

$$\rho \ddot{U}_j = T_{ij,j} \quad (3)$$

$$i, j, k, l = x, y, z$$

where T_{ij} , S_{kl} , E_k , D_i , C_{ijkl} , e_{ijk} , ϵ_{ik} and ρ correspond to stress tensor, strain tensor, electric field, electric displacement, stiffness tensor, piezoelectric stress tensor, permittivity tensor and the mass density of LiNbO₃ respectively. With a quasi-static approximation for the electromagnetic part of the problem, a scalar potential ϕ is introduced which satisfies Gauss Law

$$D_{i,j} = 0 \quad (4)$$

For the time dependence $e^{-i\omega t}$ and the appropriate mechanical and electrical boundary conditions on an ungrooved surface at $y=0$, the incident plane wave solutions for particle displacements U_{yi} and U_{zi} and static electric potential ϕ_i satisfying equations (1)-(4) may be cast in the following form;

*The work was supported by the National Science Foundation under Grant NSF-ENG-7504391.

$$\begin{aligned}
U_{y1} &= \alpha_1 \sum_{j=1}^3 (m_j e^{q_j k_s y}) e^{i(k_s z - \omega t)} \\
U_{z1} &= \alpha_1 \sum_{j=1}^3 (m_j p_j e^{q_j k_s y}) e^{i(k_s z - \omega t)} \\
\phi_1 &= \alpha_1 \sum_{j=1}^3 (m_j \gamma_j e^{q_j k_s y}) e^{i(k_s z - \omega t)}
\end{aligned} \quad (5)$$

where q_j 's are complex having positive real parts, $k_s = \omega/v_s$, and v_s is the surface wave velocity, m_j , p_j , and γ_j are known constants, and α_1 is the amplitude of the incident wave. The amplitude distributions of particle displacements and electric field intensities are shown in figures 2 and 3.

Scattering from a Single Groove

Groove Geometry

As shown in figure 1, the grooved surface is specified by

$$S(y, z) = y - \lambda \epsilon f(z) = 0 \quad (6)$$

where λ is the incident Rayleigh wavelength, $f(z)$ is chosen to be an even function of z with a maximum negative value $f(0) = -1$, and ϵ is a small number $\epsilon \ll 1$ so that the maximum groove depth $h = \epsilon \lambda$ is small compared to λ . The groove, with a width approximately equal to $2a$, is not precisely a rectangle. The deviation from a perfect rectangular shape is measured by the parameter λ .

Boundary Perturbation Formulation

A generalized Rayleigh wave given by Eq. (5) is incident on the groove specified by Eq. (6). On the grooved surface, the incident wave does not satisfy the boundary conditions. As a result, additional solutions are introduced with the total fields represented by

$$\begin{aligned}
U_y &= U_{y1} + \sum_{m=1}^{\infty} \epsilon^m U_{ym} \\
U_z &= U_{z1} + \sum_{m=1}^{\infty} \epsilon^m U_{zm} \\
\phi &= \phi_1 + \sum_{m=1}^{\infty} \epsilon^m \phi_m
\end{aligned} \quad (7)$$

which satisfy Eqs. (1)-(4). Since the expressions in Eq. (7) are power series expansions in the small parameter ϵ together with the fact that Eqs. (1)-(4) are linear, it follows that U_{ym} , U_{zm} and ϕ_m must satisfy Eqs. (1)-(4) for each m . The required boundary conditions for the solution of each order m may be obtained by series expansions in y about $y=0$ of the exact electrical and mechanical boundary conditions on grooved surface $S(y, z) = 0$. This procedure is equivalent to replacing the original problem with an equivalent problem of finding each U_{ym} , U_{zm} and ϕ_m in an ungrooved half-space which are excited by a surface distribution of induced mechanical stress sources and electrical charges located on the surface $y=0$. These surface sources represent the interaction of the incident wave with the groove. For the zeroth order solution, boundary conditions for the incident wave are reproduced from this process. For the first order, boundary conditions are obtained in the following form;

$$\begin{aligned}
T_{yy} &= S_1(0, z) \\
T_{yz} &= S_2(0, z) \\
D_{y2} - D_{y1} &= S_3(0, z)
\end{aligned} \quad (8)$$

where S_1 , S_2 and S_3 are the equivalent mechanical stress sources and induced electric charge density on the ungrooved surface $y=0$ which generate U_{y1} , U_{z1} and ϕ_1 . These sources depend on the incident wave, geometry of the groove profile and material parameters of the substrate. The first order solution expressed in its Fourier transform form may be shown to be given by

$$\begin{aligned}
U_{y1}(y, k) &= \sum_{j=1}^3 \alpha_j(k) e^{q_j k y} \\
U_{z1}(y, k) &= \sum_{j=1}^3 p_j \alpha_j(k) e^{q_j k y} \\
\phi_1(y, k) &= \sum_{j=1}^3 r_j \alpha_j(k) e^{q_j k y}
\end{aligned} \quad (9)$$

where q_j , p_j , and r_j are defined the same way as for the incident wave.

On the ungrooved surface, the boundary conditions in Fourier transform domain yields the following vector equation:

$$\vec{H} \cdot \vec{\alpha} = \frac{1}{k} \vec{S}$$

where \vec{H} is a 3×3 matrix, $\vec{\alpha} = [\alpha_1, \alpha_2, \alpha_3]^T$, $\vec{S} = [S_1, S_2, S_3]^T$ and T represents transpose.

Solving for α_1 , α_2 and α_3 , we get

$$\alpha_m(k) = \frac{1}{H} \sum_{j=1}^3 H_{jm}(S_j/k), \quad m=1, 2, 3$$

where H is the determinant of \vec{H} and H_{jm} indicates the cofactor of H at j th row and m th column.

Applying inverse Fourier transform, the first order solution may be obtained:

$$\begin{aligned}
\epsilon U_{y1} &= \frac{\epsilon}{2\pi} \int_{-\infty}^{\infty} \frac{1}{H} \left[\sum_{j=1}^3 H_{jm} e^{q_j k y} (S_j/k) \right] e^{ikz} dk \\
\epsilon U_{z1} &= \frac{\epsilon}{2\pi} \int_{-\infty}^{\infty} \frac{1}{H} \left[\sum_{j=1}^3 p_j H_{jm} e^{q_j k y} (S_j/k) \right] e^{ikz} dk \\
\epsilon \phi_1 &= \frac{\epsilon}{2\pi} \int_{-\infty}^{\infty} \frac{1}{H} \left[\sum_{j=1}^3 r_j H_{jm} e^{q_j k y} (S_j/k) \right] e^{ikz} dk
\end{aligned} \quad (10)$$

The integration in (10) is performed along the real axis of the complex k -plane. With appropriate indentations around poles and branch points, the integrals in Eq. (10) may be decomposed into branch cut contributions and residue contributions. The branch cut integrals may be identified with bulk waves generated by the groove; while the residue contributions at the

real poles $k = \pm k_s$ are identified with the scattered generalized Rayleigh waves.

The Reflected Rayleigh Wave and the Reflection Coefficient

The poles in the integrals of Eq. (10) are just the zeros of the determinant H , while $H=0$ is also the dispersion relation for the incident Rayleigh wave. The contribution from the pole at $k = -k_s$ yields the following reflected generalized Rayleigh wave

$$\begin{aligned} \epsilon U_{y1}^S &= \frac{i\epsilon}{k_s \left(\frac{dH}{dk}\right)_{-k_s}} \sum_{j,m=1}^3 H_{jm} e^{i q_m k_s y} S_j(o, -k_s) e^{-i(k_s z + \omega t)} \\ \epsilon U_{z1}^S &= \frac{i\epsilon}{k_s \left(\frac{dH}{dk}\right)_{-k_s}} \sum_{j,m=1}^3 P_m H_{jm} e^{i q_m k_s y} S_j(o, -k_s) e^{-i(k_s z + \omega t)} \\ \epsilon \phi_1^S &= \frac{i\epsilon}{k_s \left(\frac{dH}{dk}\right)_{-k_s}} \sum_{j,m=1}^3 r_m H_{jm} e^{i q_m k_s y} S_j(o, -k_s) e^{-i(k_s z + \omega t)} \end{aligned} \quad (11)$$

where $S_j(o, -k_s)$ are in terms of the incident wave and the Fourier transform of $f(z)$ and

$\frac{df}{dz}$. The transverse distributions of intensities of particle displacements and electric fields are the same as incident waves which are shown in figure 2 and 3. Therefore, the reflection coefficient R may then be defined as the absolute value of the ratio of the amplitudes of the reflected and incident Rayleigh waves at $y=0$ and $z=0$ i.e.,

$$R = \left| \frac{\epsilon U_{y1}^S(o, 0)}{\epsilon U_{y1}^I(o, 0)} \right| = \left| \frac{\epsilon U_{z1}^S(o, 0)}{\epsilon U_{z1}^I(o, 0)} \right| = \left| \frac{\epsilon \phi_1^S(o, 0)}{\epsilon \phi_1^I(o, 0)} \right|$$

Using Eq.(5), Eq. (11), it can be shown that the reflection coefficient is given by

$$R = \left| \frac{i \sum_{j,m=1}^3 H_{jm} S_j(o, -k_s)}{k_s \left(\frac{dH}{dk}\right)_{-k_s} \sum_{j=1}^3 m_j} \right| \quad (12)$$

Numerical Results and Discussions

The following material parameters for LiNbO_3 are used for computations of the reflection coefficient R given by equation (12); mass density $\rho = 4700 \text{ Kg/m}^3$, Rayleigh wave velocity $v_s = 3488 \text{ m/sec}$ and numerical values given in reference [9] are used for c_{ijkl} , e_{ijk} and ϵ_{ij} . To calculate R , the value

$\left(\frac{dH}{dk}\right)_{-k_s}$ is obtained first through the relation

$$\left(\frac{dH}{dk}\right)_{-k_s} = -\frac{v_s}{k_s} \left(\frac{dH}{dv}\right)_{-v_s}, \text{ while } \left(\frac{dH}{dv}\right)_{-v_s} \text{ is computed numeri-}$$

cally by the limiting procedure $\Delta H/\Delta V$ in the process of determining the surface wave velocity. It is found that the ratio of ΔH to ΔV converges numerically to the value equal to 28.44×10^{-3} . For a rectangular-like groove with $k \ll 1$, numerical computations result in a reflection coefficient in the form

$R = 0.6 \left(\frac{h}{\lambda}\right) \sin 2k a$. This expression implies a reflection coefficient for a single step to be $0.3 \left(\frac{h}{\lambda}\right)$ which agrees with the normal-incidence experiments using LiNbO_3 reported by Williamson and Smith [8].

The purpose of this study is to establish a theory so that reflection coefficient from a groove reflector may be calculated in terms of the material parameters of a piezoelectric substrate and the geometrical profile of the groove. The technique is by no means limited to LiNbO_3 . Calculations for the reflection coefficient using other piezoelectric substrates are in progress.

References

1. H.S. Tuan and R.C.M. Li, J. Acoust. Soc. Am. 55, 1212(1974)
2. H.S. Tuan and J.P. Parekh, in Proceedings of the 1975 IEEE Ultrasonics Symposium (IEEE, New York, 1975)
3. A. Ronnekleiv and J. Sonquet, in Ref. 2
4. R.C.M. Li, in Proceedings of the 1972 IEEE Ultrasonics Symposium (IEEE, New York, 1972)
5. G.A. Coquin and H.F. Tiersten, in J. Acoust. Soc. Am. 41, 921 (1967)
6. C.C. Tseng and R.M. White, in J. Appl. Phys., 38, 4274, Oct. (1967)
7. J.J. Campbell and W.R. Jones, in IEEE Trans. Sonics and Ultrasonics, 15, 209, Oct. (1968)
8. R.C. Williamson and H.I. Smith, in IEEE Trans. Microwave Th. Tech., 21, 195, April (1973)
9. A.W. Warner, et.al., J. Acoust. Soc. Am. 42 Oct. (1966).

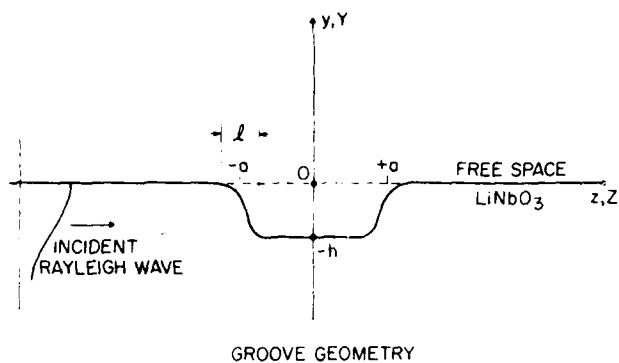


Figure 1. Geometry of the problem

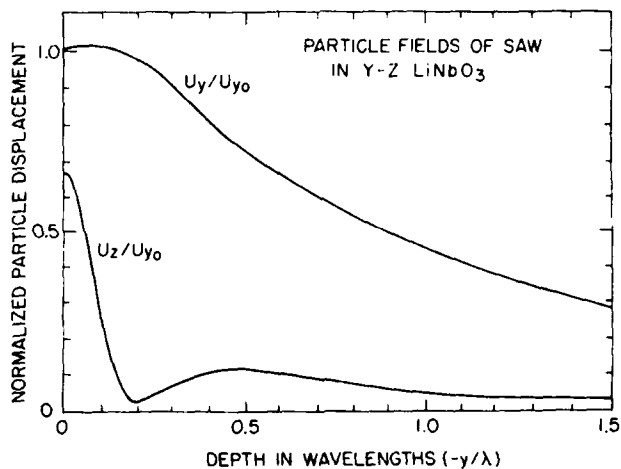


Figure 2. Normalized distribution of particle displacements U_y and U_z for incident and reflected Rayleigh waves

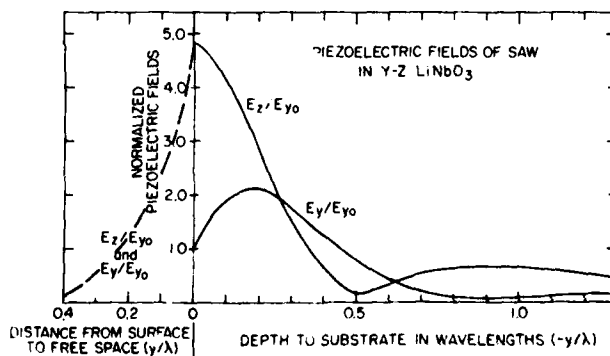


Figure 3. Normalized distributions of electric field intensities E_z and E_y for incident and reflected Rayleigh waves

UHF RANGE SAW FILTERS USING GROUP-TYPE UNI-DIRECTIONAL INTERDIGITAL TRANSDUCERS

J.Otomo, S.Nishiyama, Y.Konno
Nihon Dempa Kogyo Co., Ltd. *

and

K. Shibayama
Research Institute of Electrical Communication
Tohoku University

Summary

In the frequency range more than 200 MHz, it has been often experienced that the fabrication of quartz crystal filters meets difficulties, because of the thickness of the crystal plate. To overcome these difficulties, we have investigated and developed Surface-Acoustic-Wave filters. In this paper we report the several experimental result of the SAW filters constructing with Group-Type Uni-directional Interdigital Transducers which has many great advantages in mass production for its simple fabrication.

Group-Type Uni-directional Interdigital Transducer were fabricated on a LiNbO_3 substrate by the processes of an Electron Beam Exposure and chemical etching. A small inductor is employed for 90° phase-shifter. Consequently, a practically applicable SAW Filter providing low insertion loss and small ripple is achieved in 400 MHz and 800 MHz frequency range with fundamental mode.

Introduction

To get low insertion loss is inconsistent with to get small passband ripple in Surface Acoustic Wave filters with conventional Bi-directional transducers [1]. This problem is caused by the second order effect based on bi-directionality of the transducer.

Following two methods have been proposed to solve this problem. One is the triple transit echo (TTE) cancellation method [2][3][4][5] typified by the three-transducers, and the other is the Uni-directional transducer [6][7]. Since the former method basically consists of Bi-directional transducers, sufficiently it cannot solve the above problem. On the other hand, there are two types of Uni-directional IDT, the three phase type using 120° or 60° phase shifter and group type using 90° phase shifter. Both of them have the characteristics of low insertion loss and small passband ripple. However, since the electrode construction of the three phase type is complicated in fabrication, it is difficult to realize the SAW filters in UHF range.

Group-Type Uni-directional transducers have recently been proposed by Yamanouchi et al. They solved the problem applying this type to the VHF range [7]. This type transducer has many great advantages in mass production because of its simple fabrication. In this paper, theoretical investigation of Group-Type Uni-directional Interdigital transducer and the experimental results of SAW filters using this transducers are reported.

Equivalent Circuit of Group-Type Uni-directional IDT

As shown in Fig. 1, two interdigital transducers S and R with number of electrodes N are arranged with distance of $(N \cdot L/2 + L/4)$ and are connected to an

electric source with 90° phase difference each other. The surface acoustic waves are propagated to the forward direction. Generally, it is difficult to get a good directionality using only one group transducer having a small number of electrodes. In order to obtain a good characteristic, it is necessary to arrange many groups collinearly so that the signals from each group can be summed in-phase. Then the electrode structure in Fig. 2(a) is called N-2 type. The one in Fig. 2(b) is called N-4 type.

The equivalent circuit of the Group-Type Uni-directional transducer can be represented as shown in Fig. 3, applying the Mason's equivalent circuit of a crossed-field model for Bi-directional IDT [8] to the transducer R and S respectively. It can be represented as a three ports network adding one electric port to the Bi-directional IDT equivalent circuit.

As well known, in the SAW filter with Bi-directional IDT, it has been pointed out that the passband ripple is caused by TTE which is reflection between the acoustic ports. So, the study of impedance mismatching at the port is very important. We have investigated the admittance Y_{out} at the acoustic port under the condition that R_{in} is connected to the electric port as shown in Fig. 3. The admittance is given by

$$Y_{out} = \frac{T_{21} \cdot R_{in} + T_{11} \cdot Z_d}{Z_d(T_{22} \cdot R_{in} + Z_d \cdot T_{12})} \quad (1)$$

$|Y_{out}|$ of 6 groups transducer (N-4-6) for various values of electric source resistance R_{in} is calculated and results are shown in Fig. 4. In computation, the electromechanical coupling coefficient K is assumed as $K^2=0.0554$, which corresponds to that of rotated $Y 128^\circ \text{LiNbO}_3$ [9].

$|Y_{out}|$ is very sensitive to change of the electrical condition, but, under optimum condition in passband, $|Y_{out}|$ is almost equal to the characteristic impedance (Z_d) of the transmitting material. The variation of $|Y_{out}|$ of the Bi-directional IDT is several times larger than that of the Group-Type Uni-directional IDT as shown in Fig. 5. $|Y_{out}|$ for various values of R_{in} was also calculated and results are shown in Fig. 5. The mismatching at the acoustic port is still large for a change of electric source resistance R_{in} . Therefore, matching by change of R_{in} cannot be expected.

From these results, if we use a Group-Type Uni-directional IDT, a small passband ripple filter characteristics can be expected.

Next, as shown in the equivalent circuit of Fig. 3, transmitting transducer TR1 and receiving transducer TR2 are connected by a transmission line having the characteristic impedance Z_d , propagation constant γ and length l . The total transfer matrix of the circuit is given by the product of the individual

F/6 9/5

UNCLASSIFIED

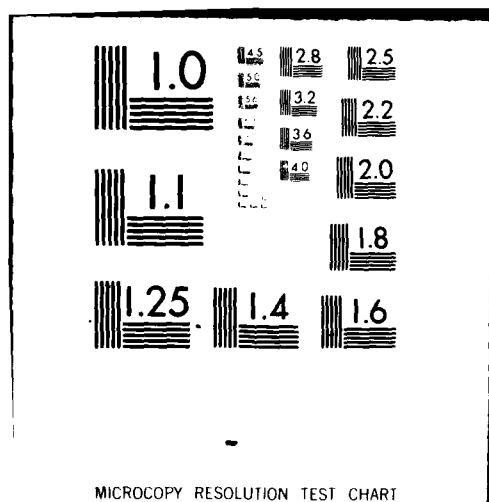
NL

4 of 7

AL: 3
CH: 3

■

This image displays a 10x10 grid of 100 squares. Each square contains a unique pattern of black dots or small shapes. The patterns are diverse, ranging from a single dot in the center to a solid black square. Some squares contain multiple dots arranged in a grid, while others have larger, more complex shapes. The overall effect is a dense, textured field of black marks on a white background.



transfer matrix.

$$\begin{pmatrix} E_1 \\ I_1 \end{pmatrix} = \begin{pmatrix} T_{11} & T_{12} \cdot Z_d \\ T_{21}/Z_d & T_{22} \end{pmatrix} \begin{pmatrix} E_2 \\ I_2 \end{pmatrix} \quad (2)$$

Therefore, the operating attenuation is given by,

$$S_D = 20 \log \left| \frac{R_{out} \cdot T_{11} + Z_d \cdot T_{12} + R_{in} \cdot R_{out} \cdot T_{21} / Z_d + R_{in} \cdot T_{22}}{2 \sqrt{R_{in} \cdot R_{out}}} \right| \quad (3)$$

where R_{in} and R_{out} are the resistance of the source and load. From equation (3), operating attenuation S_D is calculated for group type N-4-6 transducer as shown in Fig. 6. In this case, R_{in} and R_{out} are $0.1785 \times 10^{-3} \cdot Z_d$ respectively, and Z_d equal to $0.28011 \times 10^3 (\Omega)$.

Experimental Results

The 400 MHz and 800 MHz range filters are experimented, where a rotated 128° Y-cut, X-propagation LiNbO_3 plates are used. Aluminium of $1500 - 3000 \text{ \AA}$ thickness are deposited and interdigital electrodes are fabricated by Electron Beam exposure and chemical etching processes. Experiment was performed by 50 Ω measurement system.

The result of experiment of the 400 MHz range are shown in Fig. 8. The transducer construction is N-4-6 type and the overlap length W is about 0.4 mm and center-to-center spacing m of the electrodes is about $2.36 \mu\text{m}$. An air cored coil with a diameter of 5 mm is used as the 90° phase shifter. As shown in this figure, a minimum insertion loss of 2.6 dB and less than 0.1 dB passband ripple were obtained. A center frequency f_0 is 423 MHz . The radiation impedance is about 50Ω . From these results, insertion loss and passband ripple of the experimental results are in good agreement with the calculation of the equivalent circuit.

For the 800 MHz range, an N-4-8 type transducer having an overlap length $W=0.1 \text{ mm}$ and $m=1.5 \mu\text{m}$ is used as the transmitting and receiving transducers. An air cored coil with a diameter of 3 mm is used as the 90° phase shifter. The results of the experiments are shown in Fig. 9. Minimum insertion loss of 5.6 dB and the passband ripple is 0.2 dB . A center frequency f_0 is 866 MHz . The radiation impedance is about 50Ω . The passband ripple is almost in agreement with the calculated results of the equivalent circuit model, but the insertion loss is about 3 dB larger, the increasing of loss would be due to the effect of the electrode resistance and propagation loss of waves.

It should be attention on the spurious response for Group-type Uni-directional IDT. The center frequency of this spurious response will approach the passband when the number of electrodes in one group are increased. For example, in the case of the N-2 type transducer, the spurious signal f_s arises at frequency $f_s/f_0=0.66, 1.33$ and, in the N-4 type transducer, the spurious signal arises at $f_s/f_0=0.8, 1.2$.

However, this spurious response can be suppressed by suitable combination of the transducer which has different number of electrodes in one group. As an example, if we choose the combination of an N-4 type transmitting transducer and N-2 type receiving transducer, the results are shown in Fig. 10. From this figure, the spurious response suppression is improved to more than 20 dB .

In order to obtain the filter characteristics with suppressed sidelobes, we combined an N-2 type

transducer weighted by the weighting function $\Pi^2 \sin x / (\Pi^2 + x^2) x$ with an unweighted N-4 type transducer. The results are shown in Fig. 11. Sidelobe suppression is 35 dB , minimum insertion loss is 5.2 dB and the passband ripple is less than 0.2 dB . The increase of insertion loss are due to mismatching at electric port.

Finally, an exterior view of a SAW filter with this transducer is shown in Fig. 12.

Conclusion

In order to obtain low insertion loss and small passband ripple filters in UHF range, Group-Type Uni-directional IDT are used. The results of experiments for the 400 MHz range are minimum insertion loss of 2.6 dB and less than 0.1 dB passband ripple. Similarly for the 800 MHz range, minimum insertion loss of 5.6 dB and less than 0.2 dB passband ripple is obtained. The calculated results using the equivalent circuit model are in good agreement with the experimental results, and the equivalent circuit is useful to analysis of Group-Type Uni-directional IDT. The structure of the 90° phase shifter is simple and miniature UHF range SAW Filters are achieved.

Acknowledgments

The authors gratefully acknowledge the guidances and advices of Assistant Professor K. Yamanouchi and Dr. H. Sato, Tohoku University for the development SAW filter devices. Also, the authors would like to thank Mr. Y. Noguchi, S. Nakamura and S. Suzuki for their assistance in experiment and calculation in this work.

Reference

- (1) H. Sato, K. Yamanouchi, K. Shibayama and S. Nishiyama "On the Design of Elastic Surface Wave Filters with no Tuning Coil", Proceeding of the 28th Annual Sympo. Freq. Cont., 1974, P.286
- (2) A.J. DeVries, J.F. Dias, J.N. Rypkema and T.J. Wojcik, "Characteristics of Surface-Wave Integratable Filters (SWIFS)", IEEE Trans., Vol. BTR-17, No.1, Feb. 1971, P.16.
- (3) M.F. Lewis, Triple-transit Suppression in Surface Acoustic-Wave Devices", Electronics Letters Vol.8, 16th Nov. 1972, P.553.
- (4) K. Nishikawa, "Suppression of Multiple-Transit-Echoes in Narrowband Acoustic Surface Wave Filters", Report of the 1974 Spring Meeting of the Acous. Soc. of Japan, June 1974, P.513.
- (5) J. Minowa, F. Ishihara and S. Yoshikawa, "Characteristics of Low loss Elastic Surface Wave Filters", Report of the 1974 Spring Meeting of the Acous. Soc. of Japan, June 1974, P.525.
- (6) R.C. Rosenferd, R.B. Brown and C.S. Hartman, "Uni-directional Acoustic Surface Wave Filters with 2 dB Insertion Loss", Ultras. Symp. Proceedings IEEE 1974 P.425.
- (7) K. Yamanouchi, F.M. Nyffeler and K. Shibayama "Low Insertion Loss Acoustic Surface Wave Filter using Group-Type Uni-directional Interdigital Transducer", Ultras. Symp. Proceedings IEEE 1975 P.425.

- (8) W.R. Smith, H.M. Gerard and W.R. Jones,
"Analysis and Design of Despersive Interdigital
Surface-Wave Transducers", IEEE Trans. MTT-20,
No.7, July 1972 P. 458.
- (9) K. Shibayama, K. Yamanouchi, H. Sato and T. Meguro,
"Optimum cut for Rotated Y-cut LiNbO₃ Crystal using
as the substrate of Acoustic-surface-wave Filters",
Proceedings IEEE May 1976 P.595.

* 21-2 Nishihara 1-chome, Shibuya-ku, Tokyo 151,
Japan

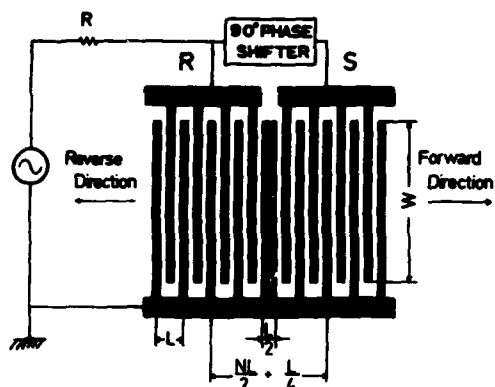


Figure 1 Group-Type Uni-directional
Interdigital Transducers with
90° phase Shifter

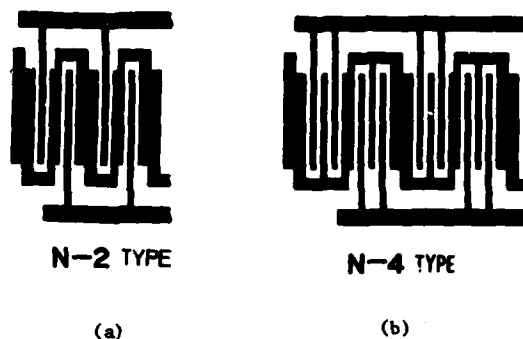


Figure 2 The Examples of Group-Type
Uni-directional IDT

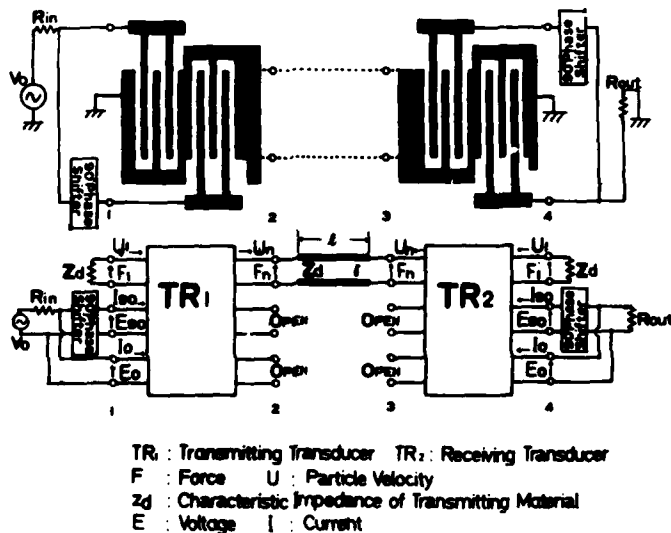


Figure 3 Schematic Diagram of Group-Type Uni-directional IDT including
Propagation Line.

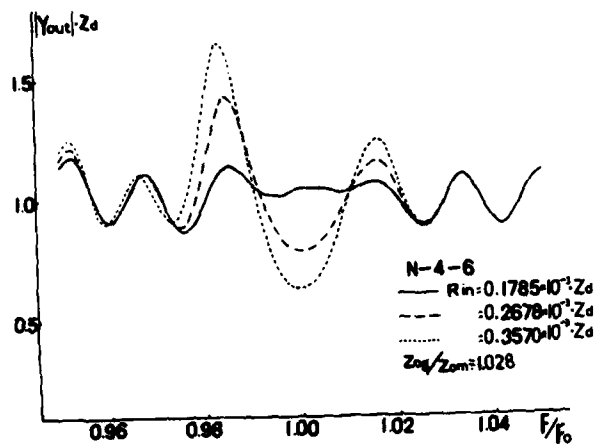


Figure 4 Admittance $|Y_{out}|$ at Acoustic Port of Group-Type Uni-directional IDT.

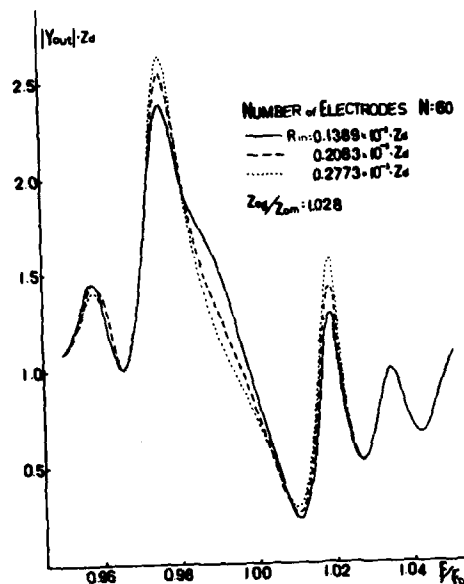


Figure 5 Admittance $|Y_{out}|$ at Acoustic Port of Bi-directional IDT.

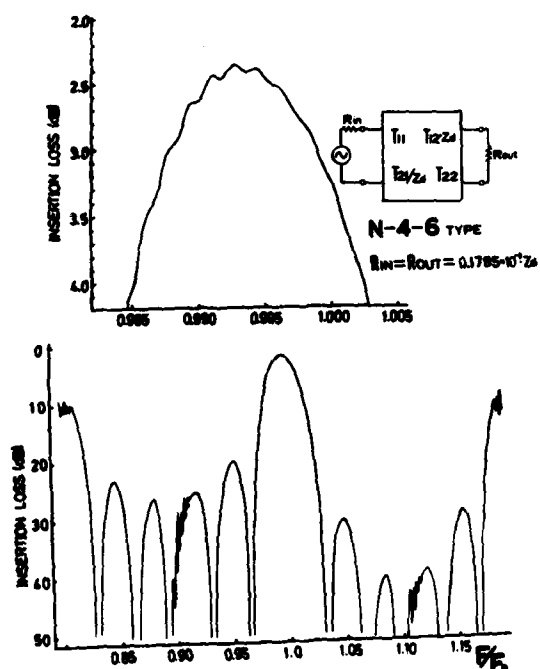


Figure 6 Calculated Operating Attenuation of Group-Type Uni-directional.

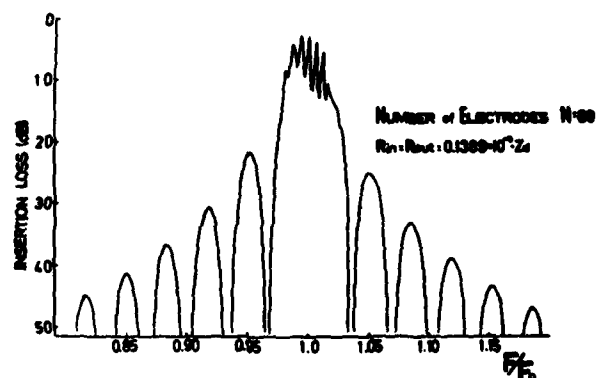


Figure 7 Calculated Operating Attenuation of Bi-directional IDT.

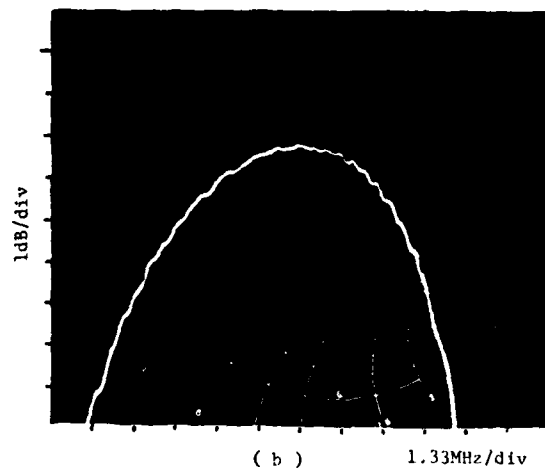
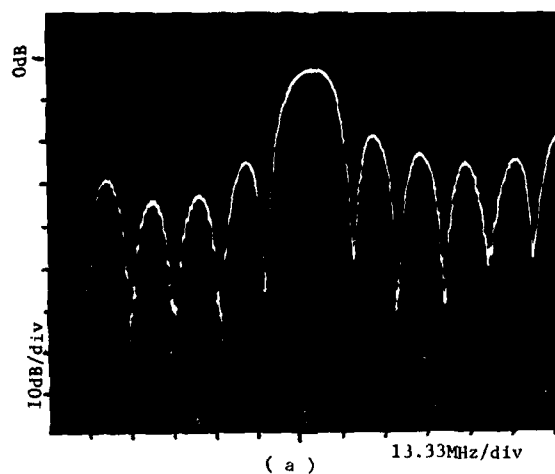


Figure 8 Measured Frequency Response of 400 MHz Range with N-4-6 Type.

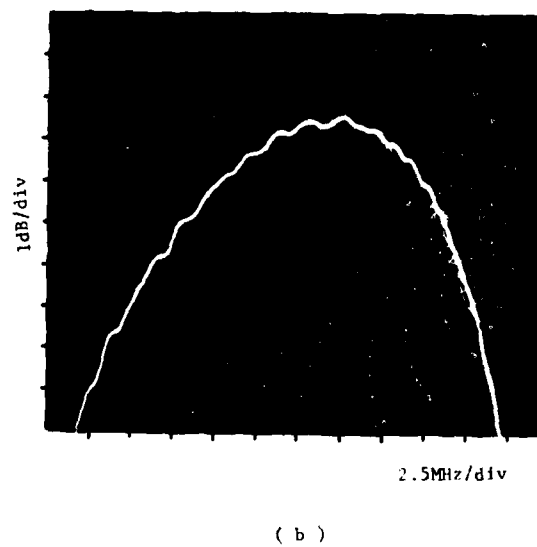
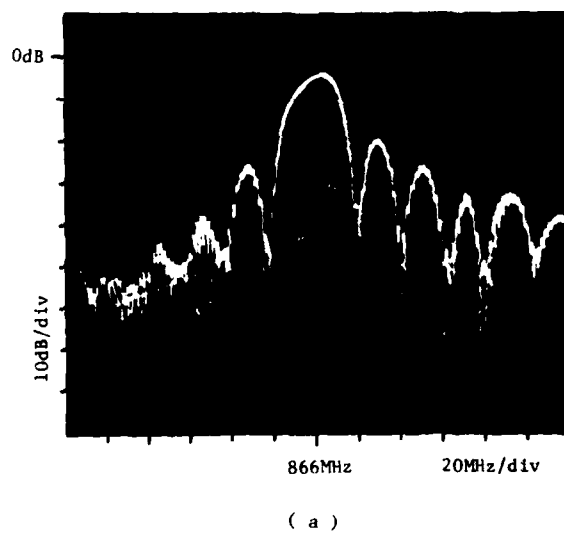
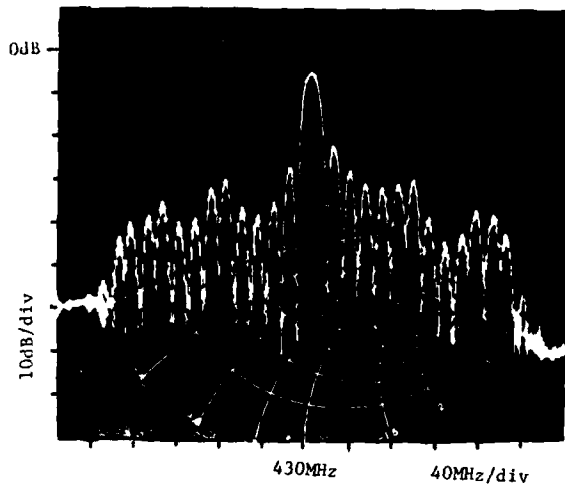
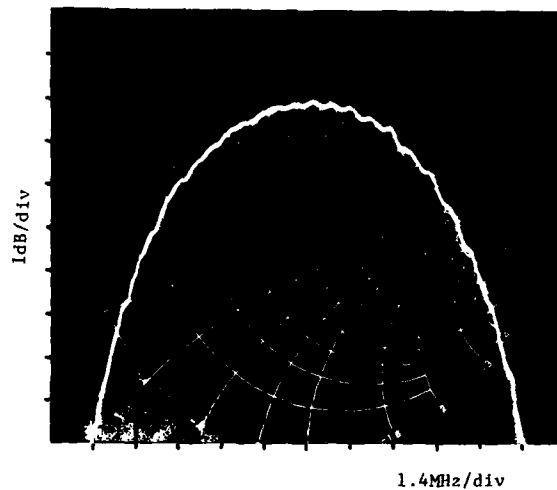


Figure 9 Measured Frequency Response of the 800 MHz Range with N-4-8 Type

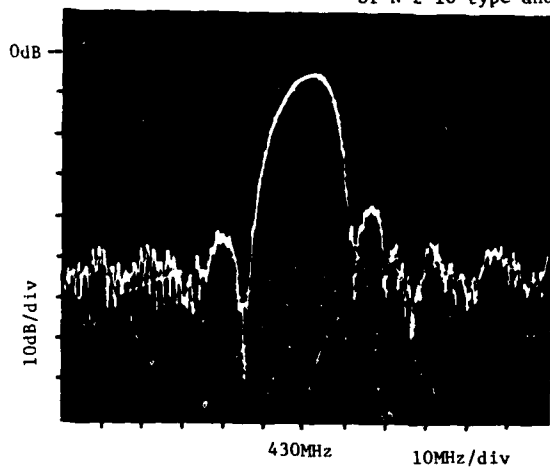


(a)

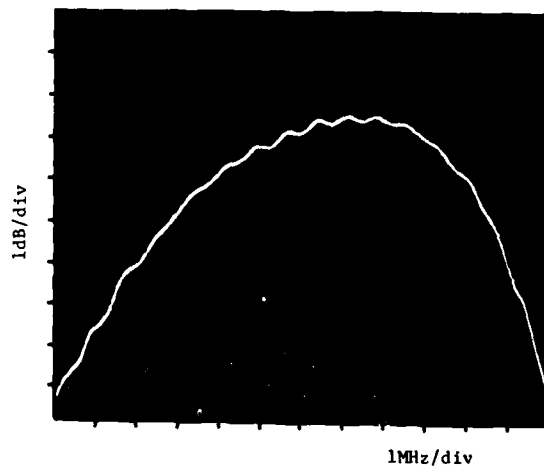


(b)

Figure 10 Measured Frequency Response of the Filter with Combination of N-2-10 type and N-4-6 Type Electrode.



(a)



(b)

Figure 11 Measured Frequency Response of the Filter with Apodised N-2 Type Electrode and Unapodised N-4 Electrode.

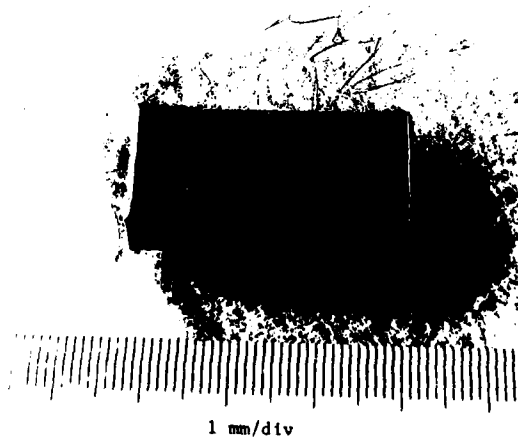


Figure 12 An Exterior View of SAW Filter Device with Group-Type Uni-directional IDT.

PROGRESS REPORT ON SURFACE ACOUSTIC WAVE DEVICE MMT - II

Alan R. Janus and LeRoy Dyal III
Hughes Aircraft Company
Fullerton, California

Summary

In July 1975, Hughes began a two-year, USAECOM sponsored Manufacturing Methods program on surface acoustic wave devices. A status report is presented which identifies the extent to which program objectives have been met in terms of design, fabrication, packaging and test considerations.

The object of the program is the establishment of a production capability for SAW devices of varied design and material. Specifically, the contract identifies six test vehicles with rigid electrical and environmental specifications. These are band-pass filters, tapped delay lines and pulse compression filters with center frequencies of 100 to 200 MHz on both lithium niobate and ST-quartz. The program is divided into four phases. During the first, design phase, the six devices were designed, fabricated and tested to specification. The second phase was directed at the redesign or design modification required by the failure of first phase samples to meet specification and resulted in a finalization of the electrical specification for the balance of the program.

The next phase consisted of testing these devices to a MIL-STD-883 type environmental stress at a high sampling rate. Results of this testing are discussed. The final phase saw the establishment of a pilot line with the production of one hundred fifty each of the device types. Capacitance probe testing is discussed as a cost-effective alternative to visual inspection at high magnification. In testing, statistical sampling is employed in order to ensure the conformance of the devices to both electrical and environmental specifications. An important element of this phase is the development of realistic manufacturing cost data on SAW devices that are representative of the technology.

Introduction

The results of work performed during the first two phases of the program have been described earlier.¹ This paper deals with the results of the third and fourth phases of the program. It encompasses the environmental testing of a broad range of surface acoustic wave device designs. It also describes the pilot line production of these devices.

Program Requirements

The overall objective of the MMT program is the establishment of a production capability for SAW device functions that are representative of a proven technology. Table I summarizes the six device designs chosen by ECOM for the program. They are broadly classified into bandpass, phase coded tapped

delay line, and pulse compression filters. Substrate materials specified are lithium niobate and ST-quartz.

In order to address the design, electrical test, environmental and volume fabrication requirements of the program in an orderly fashion, the program is broken into four phases as shown in Table II. During the first phase, the six devices were designed, fabricated and tested to specification. Ten devices of each of the six designs were delivered to ECOM at the end of this phase. In order to meet the program requirements for auto correlation (phase coded tapped delay line filters) and pulse compression line test, the appropriate reverse coded tapped delay lines and pulse expansion lines were included in this shipment.

During the second phase of the program, testing and analysis of those devices that failed to meet full specification was performed. This redesign effort resulted in a finalization of electrical specification for the balance of the program. Ten devices of each of the six designs, fully adherent to the finalized specification, were delivered at the end of this phase.

In the third phase of the program, environmental testing was performed on the devices to a MIL-STD 883B type of requirement at a high sampling rate. The completion of this phase resulted in the delivery of fifty each of the six device types.

During the final phase of the program, a pilot line was established for the production of one hundred fifty of each of the device designs, fully adherent to the electrical and environmental requirements of the program. An important element of this phase was the evaluation of capacitance probe testing as a cost effective alternative to visual inspection at high magnification. Manufacturing cost data was defined for each of the device types.

Program Accomplishments

Phase III

Test Results - Device designs finalized during the Phase II effort were fabricated in accordance with the process flow defined in Table III. These devices were then subjected to various of the environmental stresses listed in Table IV. Test allocation, sample size for each lot of fifty devices and accept/reject criteria are summarized in Table V.

In all cases, the number of device failures did not exceed the limits set forth by the program. No pattern of failure attributable to environmental stress could be discerned. Failures detected at final electrical test were random. The cause is most likely process related.

Phase IV

Capacitance Probe Test - Visual Inspection for open and short defects at 250X magnification of the wafer shown in Figure 1 typically consumes one and one-half hours of labor. In analogy with integrated circuit probe testing, a capacitance probe test has been evaluated as a cost effective substitute for visual inspection (Figure 2A). Prior experience on a limited number of samples using manual measurement (Figure 2B) showed the technique to be a viable one. The purpose of the MMT evaluation was to (1) determine the efficiency of a semiautomated approach (Figure 2C) and (2) to establish a nominal value of capacitance for open defects on a given design, based on final electrical data over a statistically valid sample.

A Fairchild Sentry 610 Integrated Circuit Probe Station was modified for capacitance measurement. It was determined that wafers of the kind shown in Figure 1 could be scanned in ten minutes yielding a computer print-out (mapping) of capacitance value for each transducer. The data indicates that short defects can be effectively screened by the technique. In fact, a yield improvement of up to five percent has been noted over visual inspection. This has been attributed to operator visual misinterpretation of (1) high resistance shorts that are electrically non-functional and (2) non-conductive debris which is mistaken for metal bridging. With regard to open defects, the final device electrical data is incomplete. It is therefore not possible to complete an assessment of the nominal capacitance value at this time.

It is felt that, with further automation, the probe time could be reduced to three minutes per wafer. The potential also exists for the addition of a controlled voltage short burn-out during this inspection, leading to further yield increase at the wafer level.

Fabrication - During the course of volume production, a number of cost versus technical trade-offs became evident.

In Figure 1, in-situ resistors on the substrate form the R of the RL tuning networks commonly used on these devices. The formation of these aluminum resistors at the same time the transducers are defined not only offers significant cost savings in labor and materials, but leads to an improvement in reliability by reducing the number of discrete components in the device.

SAW devices are materials cost intensive. Every effort is therefore made at the design level to maximize the number of die per wafer. This is, however, limited by dicing considerations.¹ Quartz requires a thicker saw blade than does lithium niobate. The kerf loss for quartz typically approaches 20 mils versus 3 mils for lithium niobate. In addition, sufficient area must be allowed for at the ends of the die for the application of acoustic absorber materials (DC-1340). Insufficient absorber contributes to the presence of spurious signals. Even the presence of occluded bubbles in the absorber can be related to spurious effects. In future work, consideration should be given to the cost effectiveness of dice taper geometries.

Program requirements for the BP-LN design imposed a metallization thickness lower limit (1000Å) as well as an insertion loss range (20 ± 1.5 dB). These constraints operated to reduce the usable insertion loss range on that device during the final phases of the program due to the fact that a new evaporator yielded lower resistivity aluminum than that utilized during the initial designs. This is shown in Figure 3.

The cost savings of substituting semiconductor pin packages for machined chassis are estimated at 10 to 1. However, this substitution can cause difficulties in achieving feedthrough suppression levels of 50 dB. Grounding through package pins alone will not suffice. Adequate grounding of the package base, both in test and a functional environment is critical. In the bandpass filter designs, tuning inductor orientation (rotational) was found to contribute to feedthrough. One solution to excessive feedthrough in the phase coded tapped delay lines was the inclusion of a ground pad between the input transducer and output tap array during photolithography. It should be pointed out that none of these problems were encountered with the use of machined chassis.

Cost - The manufacturing cost of the six device types, packaged and tested, is estimated to range from 20 to \$50 in quantities of 200. Of the pulse compression and bandpass filters, the quartz devices range higher; approximately \$30. The phase coded tapped delay line units average \$50, reflecting the impact of large die size, higher unit area material cost, layout expansion due to sawing requirements and increased package size.

Overall, package cost is in the range of 15 to 25% of manufacturing cost. Future effort in this area should address the requirement of still lower cost packaging. The impact of higher volume production on cost reduction should not be overlooked. The estimated manufacturing cost of the BP-Q device in quantities of 1000 is \$20. The cost of tuning elements required to achieve low VSWR, especially in regard to component attach labor, is high. In view of the relative insignificance of this parameter at IF, the imposition of this requirement should be closely examined at the design level.

Conclusions

Fabrication and test efforts during the final phases of a surface acoustic wave MMT program have shown that a variety of SAW device designs, capable of withstanding MIL-STD 883B type environmental requirements, can be produced in relatively small quantities at low cost. In the area of future cost reduction, consideration should be given to packaging and the sophistication of tuning requirements.

Acknowledgements

The authors wish to express their appreciation to T. W. Bristol for his encouragement and to G. Blurton for device fabrication. This work was performed under USAECOM Contract No. DAAB07-75C-0044.

References

1. Janus, A.R., Proc. 30th Annual Frequency Control Symp., Atlantic City, NJ, 157 (1976)

BANDPASS FILTERS	BP-Q	100 MHz	ST-QUARTZ
	BP-LN	150 MHz	Y-Z LITHIUM NIOBATE
BIPHASE CODED TAPPED DELAY LINE FILTERS	TDL-100	100 MHz	ST-QUARTZ
	TDL-200	200 MHz	ST-QUARTZ
PULSE COMPRESSION FILTERS	PC-Q	150 MHz	ST-QUARTZ
	PC-LN	150 MHz	Y-Z LITHIUM NIOBATE

I SUMMARY OF MMT DEVICE DESIGNS

PART/TYPE NO. _____ APPROVAL DEPT. DATE
 FLOW SHEET REV. NO. _____ ASBY. _____
 CHARGE NO. _____ TEST _____
 TEST SPEC. NO. _____ PGM. MGR. _____
 Q.A. _____

STEP	OPERATION	PEI REF. NO.	BY	DATE
1.	CRYSTAL MATERIAL INSPECTION	7.18.01A		
2.	CLEAN	7.18.14		
3.	METALLIZATION DEPOSITION	7.18.16		
4.	RESIST COAT	7.18.17		
5.	EXPOSE	7.18.18		
6.	DEVELOP	7.18.19		
7.	ETCH	7.18.20		
8.	OPTICAL INSPECTION	7.18.08A		
9.	PROTECTIVE COAT (RESIST)	7.18.17		
10.	DICE WAFER	7.18.22		
11.	TOROID ATTACH	7.18.23		
12.	CRYSTAL ATTACH	7.18.24		
13.	BOND CRYSTAL (T.C. BOND)	7.18.25		
14.	TUNE	7.18.26		
15.	O.C. VISUAL, PRE-CAP (METHOD 2017)	QMS-M-022, Q23, Q24, Q25		
		N/C		
16.	SEAL	7.18.11B		
17.	MARK COVER	7.18.02A		
18.	GROSS LEAK (METHOD 1014)	7.18.08B		
19.	FINE LEAK	7.18.10C		
20.	FINAL ELECTRICAL TEST	7.18.26		
21.	O.C. VISUAL FINAL (METHOD 2009)	QMS-M-028		
		N/C		

III SAW PROCESS FLOW SHEET

DEVICE TYPE	GROUP TEST (EA. TYPE)	SAMPLE SIZE (EA. TYPE)	MAX. DETECTIVE
BP-Q, BP-LN, PC-Q, PC-LN	I	50	0
	II	12	0
	III	18	0
	IV	18	1
	V	12	0
	VI	24	2
TDL-100, TDL-200	I	50	0
	II	8	0
	III	9	1
	IV	9	1
	V	8	0
	VI	12	1

V CONFIRMATORY SAMPLE TEST PLAN

STEP I - PHASE I - FIRST ENGINEERING SAMPLES

- DESIGN SIX SAW DEVICES TO SPECIFICATION
- FAB. PACKAGE AND TEST DEVICE DESIGN
- DELIVER TEN EACH

PHASE II - SECOND ENGINEERING SAMPLES

- REDESIGN DEVICES AS NECESSARY
- FINALIZE ELECTRICAL SPECIFICATION
- DELIVER TEN EACH

PHASE III - CONFORMAL SAMPLES

- TEST DEVICES TO ENVIRONMENTAL SPECIFICATION
- FINALIZE ENVIRONMENTAL SPECIFICATION
- DELIVER FIFTY EACH

PHASE IV - PILOT RUN SAMPLES

- FINALIZE LINE
- ESTABLISH DEVICE COST
- DELIVER 100 EACH

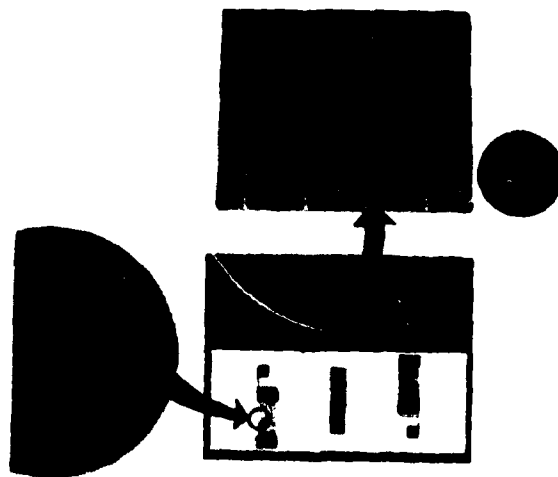
STEP II

- DEVELOP PLAN FOR MBRND. PRODUCTION RATE

II JOB DESCRIPTION AND TASKS

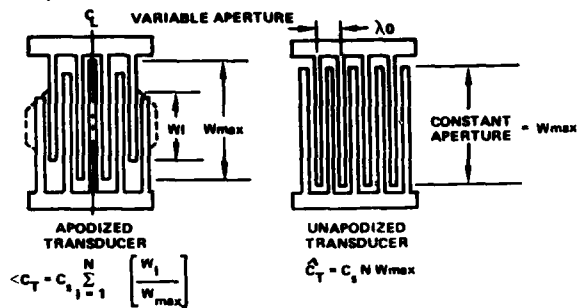
GROUP	TEST	MIL-STD	METHOD	CONDITION
I	WIRE BOND	353	2011	D
	1ST ELECTRICAL	-	-	-
II	SOLDERABILITY	883	2003	-
III	HIGH TEMPERATURE STORAGE	883	1008	A
	CENTER FREQUENCY INSERTION LOSS	-	-	-
IV	LIFE	202	108	-
	FINAL ELECTRICAL	-	-	-
V	HERMETICITY	202	1128	C
	SHORT CIRCUIT	-	-	A
VI	VIBRATION	202	201	-
	SHORT CIRCUIT	-	-	-
	SHOCK	202	212	I
	SHORT CIRCUIT	-	-	-
	THERMAL SHOCK	202	107	A
	SHORT CIRCUIT	-	-	-
	MOISTURE RESISTANCE	202	1080	-
	FINAL ELECTRICAL	-	-	-

IV CONFIRMATORY SAMPLE TEST REQUIREMENTS

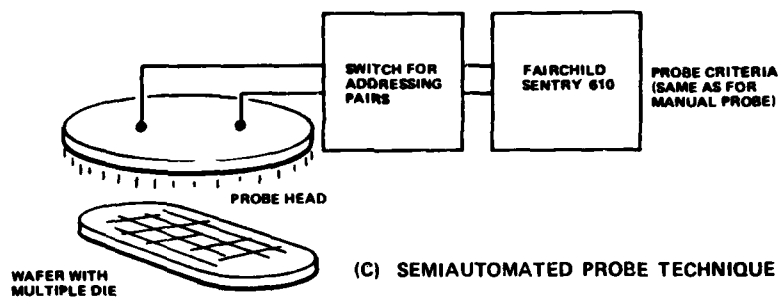
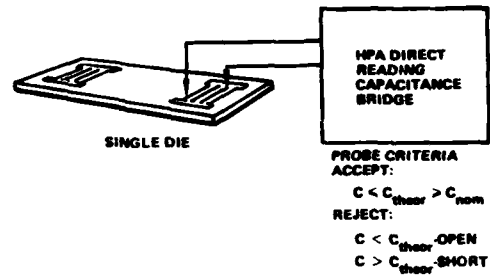


1. PROCESSING DETAIL ON BP-LN

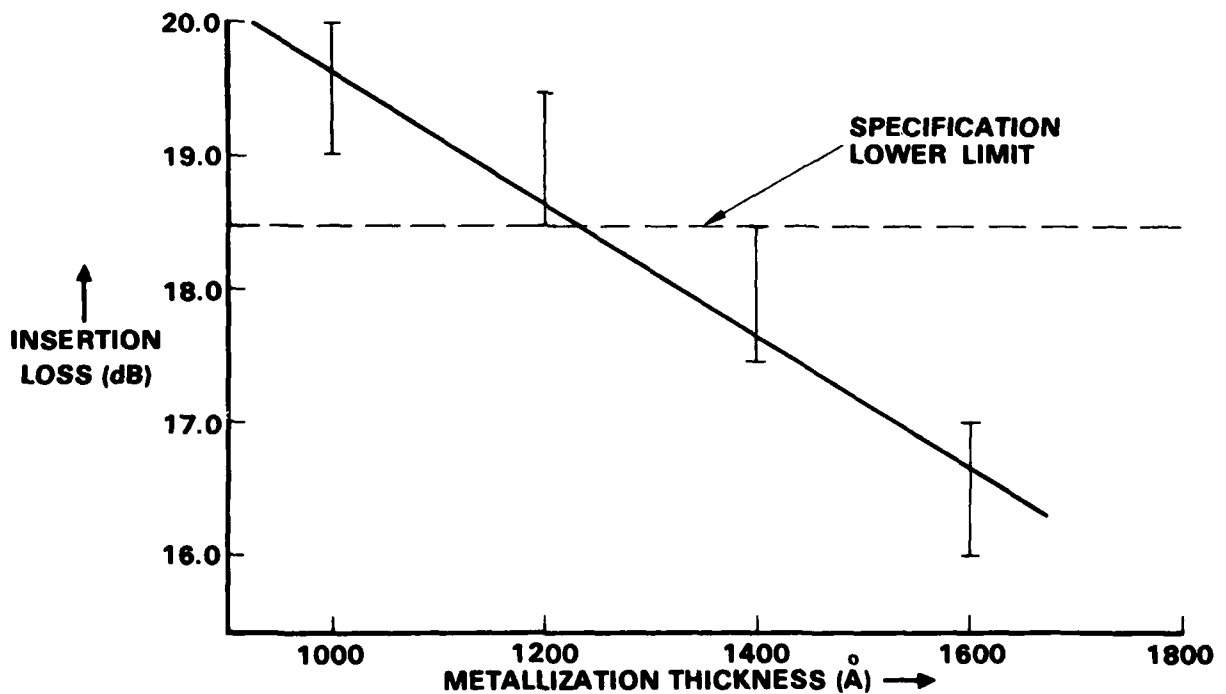
(A) THEORETICAL ORIGIN OF CAPACITANCE MEASUREMENT,
(TYPICAL ELECTRICAL PARAMETERS)



(B) MANUAL PROBE TECHNIQUE



2. CAPACITANCE PROBE TEST



3. INSERTION LOSS AS A FUNCTION OF METALLIZATION THICKNESS

THE VERSATILITY OF THE "IN-LINE" SAW CHIRP FILTER

W. J. Skudera, Jr.
US Army Electronics Technology & Devices Laboratory (ECOM)
Fort Monmouth, New Jersey

Summary

SAW chirp filters have recently been used to perform sophisticated functions such as continuous variable delay and chirp transformation. This paper will briefly cover several interesting applications of the "in-line" chirp filter used in combination with standard mixers and/or standard amplifiers. These applications include a variable delay line, a "time-ordered" multiplexer, and an oscillator. The variable delay line, which requires two standard mixers along with two chirp filters, will be discussed using relatively wide and narrow bandwidth chirp filters (DDL); and a simple cascade scheme will be shown, whereby, the less complex "in-line" filter can yield variable delays approaching those of the RAC variable device. The inherent fixed delay portion, important in certain applications, will also be considered for each type of filter.

Again, using a mixer and two "in-line" chirp filters, a "time-ordered" multiplexer can be constructed based upon the chirp transformation technique. The multiplexer scheme will be shown utilizing two identical filters instead of the normally used "up" and "down," offset frequency chirp filters. The advantage and disadvantage of using identical "in-line" SAW filters will be discussed in terms of filter fabrication ease and transform bandwidth restrictions, respectively.

The final application covered will be the chirp "in-line" SAW filter used as an oscillator; results for two different bandwidth filters will be given in terms of spectral purity, power output, and tuning ability. In addition, the tuning ability of the chirp oscillator will be compared to that of a standard delay line SAW oscillator.

Introduction

Surface acoustic wave (SAW) linear FM (chirp) filters have been used primarily in pulse compression applications to improve the resolution of a radar system. Two basic SAW chirp filters have evolved for this application: the "in-line" array¹ and the reflective array compressor (RAC) filter.² The RAC device is used, usually where large time-bandwidth products with especially large dispersions, are required; the less costly "in-line" filter is used for modest time-bandwidth products.

Figure 1 shows the difference between the two filters. The "in-line" filter uses metal electrodes, in a graded periodicity fashion, to yield a linear dispersion with frequency, and it has a fixed delay determined by the separation of the input and output transducers. The RAC filter uses etched gratings to create the linear dispersion along with non-dispersive metal input-output transducers, and it has a fixed delay determined by the spaces between the gratings as well as from the transducers to the gratings. The "in-line" fixed delay is much less than that of the RAC's and can be made as small as approximately 0.25 microsecond compared to several microseconds for a RAC.³

Figure 2 shows one of the "in-line" filters that were used along with standard mixers and/or amplifiers to perform three interesting functions. The filters were designed and fabricated as previously discussed,⁴ and were centered at 70 MHz, with two different

bandwidths (10 MHz & 2.5 MHz) and approximately 8 microseconds of dispersions, using a total of 1354 electrodes per transducer on ST-X quartz. The electrode patterns are not visible in the photo of Figure 2, but the pads for the input and output transducers are clearly seen. The response for the 10 MHz bandwidth filter is given in Figure 3 showing a passband containing the usual Fresnel ripple⁵ along with a few dB slope which it is assumed is attributed to the use of identical cosine-squared weighted transducers.

Two recent applications for SAW chirp filters, the variable delay line^{6,7} and the chirp transform^{8,9} will be briefly reviewed, and results will be presented for slight modifications to each. In addition, oscillator results will be given, using SAW chirp filters in a standard feedback loop.

Variable Delay Line

The schematic for a continuously variable delay line using SAW chirp filters (DDL-1 & DDL-2) is shown in the top portion of Figure 4. Two standard mixers are required along with a control oscillator (VCO) to make it functional. The lower portion of Figure 4 shows a series of delayed output pulses, from a narrow input pulse, whose delay is determined by the VCO setting. Reference 6 gives the details of operation and is summarized below. When the VCO is held at a constant value and the input frequency is varied, then the delay vs input frequency slope, after the first chirp filter, is an up slope, assuming DDL-1 is an up chirp. After the second mixer, spectral inversion takes place and the slope is a down slope, then passing that signal through a second up chirp filter (DDL-2) results in a constant delay at the output with an output frequency that tracks the input frequency. However, if the input frequency is held constant and the VCO is varied, then the frequency after the first mixer increases as the VCO frequency increases, while the delay after the second mixer and the output increases with respect to the increasing VCO. The frequency after the second mixer and second chirp filter (DDL-2) is compensated by the VCO, and thus, the output is at the same constant frequency as the input, but with a VCO controlled variable delay. Figure 5 shows the output results for a variable delay line using two 10 MHz chirp filters as well as two 2.5 MHz chirp filters. The control frequency vs delay slopes are 1.536 MHz/ μ s and 0.334 MHz/ μ s respectively, for a constant 70 MHz input pulse. The output delay is also shown to be constant at 12 microseconds for a constant 138 MHz VCO frequency, while the input frequency is varied across the bandpass of each chirp filter pair, i.e., from 66 MHz to 74 MHz for the wider bandwidth filters and 69 MHz to 71 MHz for the narrower bandwidth filters. The total variable delay at the midband frequency was about 8 microseconds (from 8 μ s to 16 μ s) with a minimum fixed delay of about 8 microseconds. The fixed delay is due mostly to the loading effect of the second chirp filter (DDL-2) at its mid delay value of about 6 microseconds, plus a small amount of fixed delay (about 2.5 μ s) due to the non-optimized transducer separation of the first chirp filter (DDL-1). The lower left side of Figure 6 shows this result plotted on a linear graph. Delay data is also shown when two additional identical chirp filters are cascaded at the locations indicated by the two arrows in the upper schematic of Figure 6. Thus, using four filters the variable delay is doubled while that

of the minimum delay is increased to about 18 microseconds mostly due to the loading effect of the last two filters (about 12 μ s) plus the fixed delay of the first two chirp filters. If eight filters were used, the variable delay could be extended to greater than 30 microseconds with a fixed delay of about 36 microseconds. It is to be noted, however, that cascading tends to magnify the errors present in each filter and becomes impractical after a certain point. The lower right side of Figure 6 shows, for comparison, the range of delay that Dolat⁶ obtained for two RAC filters with a minimum delay of 54 microseconds. However, the RAC variable delay operated over a wider input bandwidth than the cascaded "in-line" device. Thus, if the bandwidth were reduced to that of the "in-line" device, then the minimum delay is estimated to be about 40 microseconds for this RAC device. Therefore, if one desires to have a variable delay line that has a minimum amount of fixed delay, required for certain repeater applications,¹⁰ and only a few microseconds of variable delay, then the "in-line" approach is worthwhile. For example, if "in-line" filters were used in a 6 μ s variable delay line then, with a transducer separation of 0.25 μ s per filter, the fixed delay for a constant input frequency would be less than 3.5 microseconds. If RAC filters with a minimum of fixed delay (about 2 microseconds) were used in the same 6 μ s device, then the minimum amount of fixed delay for the device would be about 7 μ s. However, if large amounts of variable delay, greater than about 20 μ s, are required, then the fixed delay will be large in both the "in-line" and RAC devices, with the RAC approach being more desirable than the cascaded "in-line" because it would have less distortion effects over a wider frequency range. Cascading of identical "in-Line" filters, however, appears to be a useful technique for variable delays up to 20 μ s in this frequency range, but with less bandwidth than the RAC approach.

Chirp Transformation

The Chirp Transformation^{8,9} schematic is shown in Figure 7 and expressed mathematically by the top expression in the same figure; i.e., the transform is the convolution of the second chirp filter (DDL-2) response, $I(T)$, with the product of the input signal, $S(T)$, and the response of the first chirp filter, $C_1(T)$, multiplied by the impulse response of the third chirp filter (DDL-3), $C_2(T)$. This expression can also be expressed by the lower equation $V(\tau)$ in Figure 7 and is the true Fourier transform if $\omega = 2\pi\tau$, where μ is the chirp slope. One can also consider an output located at point 1 in this figure given by $R(\tau)$, i.e., eliminate the second mixer and DDL-3 filter. Comparing the $R(\tau)$ equations in Figure 7 one sees that only the phase term ($e^{j\mu\tau^2}$) is the difference between the two expressions. Thus, if just transform amplitude data is required, then $R(\tau)$ is sufficient and the DDL-3 filter plus mixer is not needed. The first two chirp filters generally are dissimilar in this processor^{8,9}; however, the following data were taken, using two identical filters. Figure 8 shows the output results displayed on an oscilloscope when two c.w. signals are simultaneously fed into the input of the processor using identical filters. Since the filters operate only over a finite time, then the c.w. signal appears as a gated R.F. pulse, and thus the transform should be a $\sin(X)/X$ response that varies linearly with the input frequency. The scope pulses in Figure 8 do resemble a $\sin(X)/X$, except some of the sidelobes are non-symmetrical, due primarily to the passband response not being perfectly flat; and the location on the time axis of the scope does vary linearly with the input frequency. Thus, since this processor separates

frequencies in the time domain, then it can be called a "time-ordered" multiplexer. Figures 8(A) and 8(C) show that two c.w. signals nominally set at 140 MHz and separated by less than 400 kHz can easily be resolved in this multiplexer independently of the chirp filter's bandwidth because resolution is dependent only upon the dispersion¹¹ of the filters. However, 8(B) and 8(D) show two c.w. signals that are separated by the maximum amount and indicate that the number of frequency bins that can be resolved at this resolution is dependent upon the filter bandwidth; i.e., for the narrow bandwidth case a maximum of only six frequencies could be simultaneously displayed, while for the wider bandwidth case about twenty five frequencies could be displayed. Figures 9 and 10 show the amplitude and compressed pulse width of the pulses of Figure 8 vs the input frequency and clearly show that the amplitude decreases while the pulse width increases toward the processor input band edges. Therefore, the most useful operation was defined as the points where the pulse width increased by 10% of its center frequency value, and this corresponded to about half of the bandwidth of each chirp filter. The exact centers of the amplitude and pulse width curves do not coincide, and again this is attributed to the slope of the filter response which could easily be corrected. However, the pulse widths are reasonably close to the theoretical values of 0.1 μ s and 0.45 μ s respectively, and are relatively constant in both amplitude and pulse width over the defined 10% pulse bandwidth. The amplitude slope of these filters could be influencing the pass band response of the processor, but this has been assumed to be a negligible effect. Thus it appears that if one requires an input bandwidth of 10 MHz for a chirp transform processor, then two identical 20 MHz filters should suffice since, in practice, the amplitude and pulse width changes are small over this range, and thus only one chirp mask, instead of two, is required to construct the processor. Another advantage of using identical filters over the dissimilar type is that one can use filters operating at half the center frequency of the input frequency by using the lower side band of the first mixer. Thus it is easier to fabricate these filters since it allows one to operate at twice the frequency of the highest operating chirp filter that can be constructed. However, the disadvantage of using the identical filter approach is that it can only be used for processor bandwidths of up to 25% since the filters themselves can only be fabricated with about 50% bandwidth.

Chirp Oscillator

Figure 11 shows a chirp filter (DDL) used as part of a feedback loop of an amplifier along with a phase shifter to form an oscillator circuit, and Figure 12 (B) shows the frequency spectrum at approximately 68 MHz when the 2.5 MHz DDL is used in the loop. When compared with a standard 60 MHz SAW oscillator, Figure 12(A), and a standard signal generator, Figure 12(C), set at 68 MHz, one sees that the spectral responses are approximately the same and that the power outputs are all about +4 dBm. The second harmonic was found to be lower, however, by about 15 dB in the signal generator over the two types of SAW oscillators, i.e., -30 dB and -15 dB, respectively.

Figure 13(A) shows the tuning range of the 2.5 MHz DDL filter when the phase shifter is changed from a nominal zero degrees to 360 degrees, covering about 250 kHz of continuous tuning ability. The 10 MHz DDL tuning range (see Figure 13(B)) is shown for a nominal (from right to left) zero, 180, and 360 degree phase shift and covers a total of about 500 kHz frequency shift. However, a standard delay line, using 19 electrode pair transducers on ST-X quartz,

whose transducer separation was equal to approximately the center frequency separation between the transducers of the DDL filters (about 1.9 cm), had only about 100 kHz per 360 degrees phase shift. Initially, one would suspect, therefore, that the bandwidth of the DDL filters influences in some way the tuning range of these filters. However, this was not the case since, when the transducer separation of the delay line was reduced to about 0.3 cm, the tuning range per approximately 360 degree phase shift (see Figure 13(D)) was about 900 kHz (the approximate 180 degree phase shift is shown in the center of the photo).

The frequency response for the 2.5 MHz DDL filter is shown in Figure 14(A). The ripples in the passband are largely from the Fresnel ripple⁵ effect for low time-bandwidth product filters. It was found, however, that the left amplitude peak was at the precise frequency that the oscillator operated at, and this corresponded to a minimum transducer separation (T_d) of about 0.65 cm. (Note: H-L and L-H represent transducers with high frequencies, H, on one end and low frequencies, L, on the opposite end). Figure 14(B) also shows a 2.5 MHz DDL (Expander) whose transducers are rotated by 180 degrees with respect to the conjugate filter of Figure 14(A); and when this filter was used as part of an oscillator circuit, it too operated at a frequency corresponding to the amplitude peak on the right and a T_d again equal to about 0.65 cm. Also, the point of operation for the 10 MHz bandwidth filters corresponded to a T_d equal to about 0.3 cm. Substituting these values of T_d into the Q equation^{12,13} given in the lower part of Figure 14, one finds that the tuning which is the reciprocal of the Q agrees reasonably well with this equation which was derived for SAW delay line oscillators. Thus, it appears that chirp oscillators operate at a passband peak and behave as delay line oscillators. In addition, chirp filters can be internally weighted¹ and one could postulate, therefore, that this peak could be optimized and perhaps even relocated using finger withdraw techniques so as to select a desired frequency in the passband.

Some initial stability measurements were also taken (see Figure 15) and show that the narrow band (conjugate filter) chirp oscillator stability was reasonably good and generally changed by less than 1.0 ppm in an uncontrolled environment over several hours of operation. Toward the end of the test the stability appeared to improve to less than 0.02 ppm over a 5-minute period. The stability of the wider bandwidth filter was somewhat less stable over the same period and may be due to a less pronounced passband amplitude peak at the frequency of operation. Further stability tests are required in order to determine the upper stability limit for this device using a more controlled environment.

Conclusion

Two recent SAW chirp filter applications have been briefly discussed with modifications made to each. The cascading of the less expensive "in-line" chirp filters was shown to be successful in increasing the variable delay of a variable delay device and the fixed delay of the filters was also discussed in terms of the delay device. Small fixed delay in the variable delay device can be obtained from chirp filters whose transducers are close together and whose dispersion is kept to a minimum. The chirp transform was shown to operate with identical filters as opposed to the usual non-identical filter approach. However, the effective input bandwidth was shown to be limited to about half the chirp filter bandwidth. In addition, the chirp "in-line" filter was used

along with a standard amplifier to form an oscillator that operates similarly to a delay line oscillator, but may have added frequency selectivity. However, further work is required to demonstrate the potential advantages of the oscillator device.

There are other applications of chirp filters such as in spread spectrum¹⁴ and frequency hopping,¹⁵ not covered in this paper, but they clearly demonstrate that the chirp filter is indeed a very useful and versatile device.

Acknowledgements

The author wishes to express his appreciation to the following ECOM personnel: Mr. G. LeMeune for doing the SAW photolithographic fabrication. Mr. D. Ruppe, Mr. G. Sommerall, and Mr. T. Bifulco for their helpful suggestions in cutting the SAW enlarged mask. Mr. N. Korolkoff and Mr. B. Farah for helpful suggestions in the mask reduction. Also, Mr. T. Lukasek of ECOM and Mr. O. Otto of the Hughes Research Laboratories for useful discussions about SAW oscillators.

References

1. W. J. Skudera, Jr., and H. M. Gerard, "Some Practical Considerations of Dispersive Surface Wave Filters," Proceedings of the 27th Annual Symposium on Frequency Control, pp. 253-261, June 1973.
2. R. C. Williamson and H. I. Smith, "The Use of Surface-Elastic Wave Reflection Gratings in Large Time-Bandwidth Pulse-Compression Filters," IEEE Trans. Microwave Theory & Tech., MTT-21, p. 195, 1973.
3. H. M. Gerard of Hughes Aircraft Co., Private Communication.
4. W. J. Skudera, Jr., and G. LeMeune, "Acoustic Surface Wave Fabrication Techniques and Results," TR ECOM-4333, August 1975.
5. C. E. Cook and M. Bernfeld, "Radar Signals," Academic Press, New York, 1967.
6. V. S. Dolat and R. C. Williamson, "A Continuously Variable Delay-Line System," 1976 Ultrasonic Symposium Proceedings, IEEE pub. #76CH 1120-SSU, pp. 419-423, Oct. 1976.
7. J. Burnsweig et al, "Electrically Controllable Time Delay," IEEE 1973 International Microwave Symposium Digest, pp. 134-136.
8. G. R. Nudd and O. W. Otto, "Chirp Signal Processing Using Acoustic Surface Wave Filters," 1975 Ultrasonics Symposium Proceedings, IEEE pub. #75 CHO 994-4SU, pp. 350-354, Sept. 1975.
9. R. M. Hays et al, "Surface Wave Transform Adaptable Processor System," 1975 Ultrasonics Symposium Proceedings, IEEE pub. #75 CHO 994-4SU, pp. 363-370, Sept. 1975.
10. D. Klose, of USAECOM, Private Communication.
11. G. R. Nudd and O. W. Otto, "Real-Time Fourier Analysis of Spread Spectrum Signals Using Surface Wave-Implemented Chirp-Z Transformation," IEEE Transactions on Microwave Theory & Techniques, pp. 54-56, Jan. 1976.

12. H. G. Vollers and L. T. Claiborne, "RF Oscillator Control Utilizing Surface Wave Delay Lines," Proceedings of the 28th Annual Symposium on Frequency Control, pp. 256-259, May 1974.
13. M. Lewis, "The Surface Acoustic Wave Oscillator - A Natural and Timely Development of the Quartz Crystal Oscillator," Proceedings of the 28th Annual Symposium on Frequency Control, pp. 304-314, May 1974.
14. K. Ikrath, W. Skudera, and C. DeSantis, "Short-Range Radio Communications System for Military Operations in Built-up Areas," TR ECOM-4481, March 1977.
15. C. Atzeni et al., "Programmable Signal Processing by Analog Chirp-Transformation Using SAW Devices," 1975 Ultrasonics Symposium Proceedings, IEEE pub. #75 CHO 994-4SU, pp. 371-376, Sept. 1975.

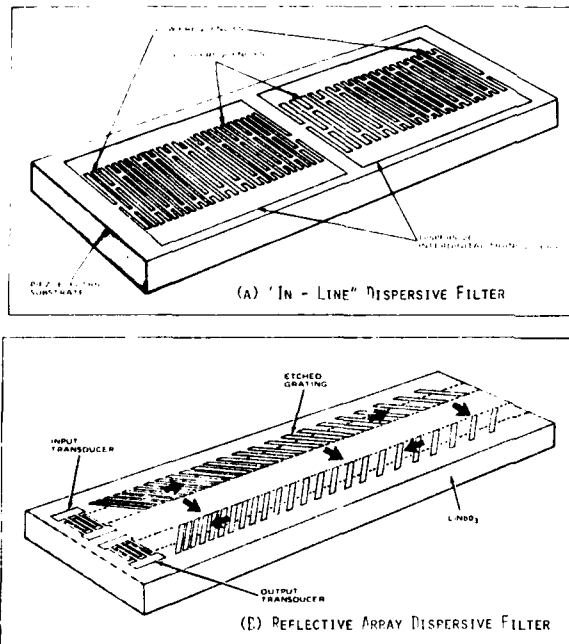


Fig. 1 Chirp Filter Schematics (A) "In-Line" Chirp (Expander) Transducer Pattern, (B) Reflective Array Dispersive Filter Showing a Down Chirp Grating Pattern.

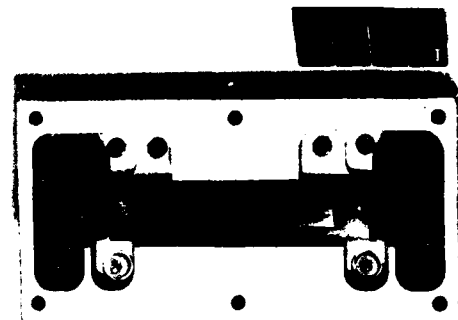


Fig. 2 Photograph of a 70 MHz "In-Line" Dispersive Filter.

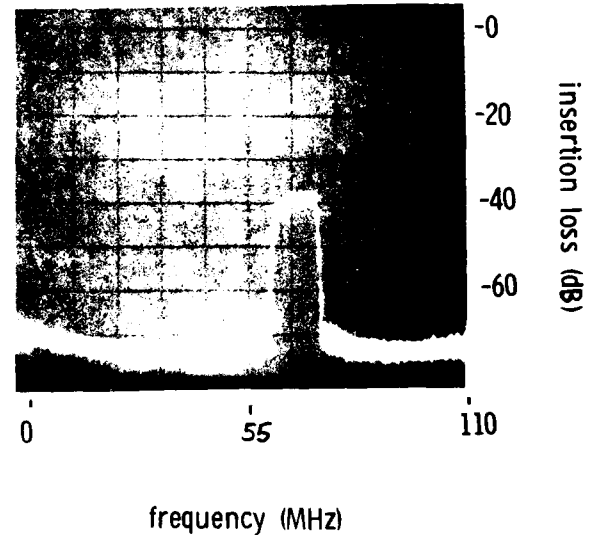


Fig. 3 Frequency Response of a 10 MHz Bandwidth "In-Line" Chirp Filter.

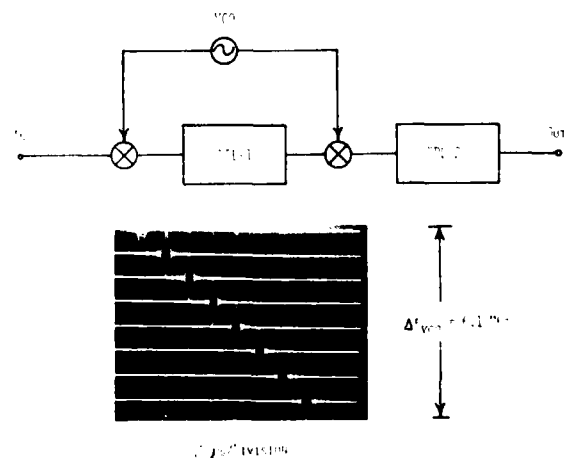


Fig. 4 Schematic of Variable Delay Line (Top) and Photograph (Bottom) Showing a Series of Delayed Output Pulses Taken at Discrete VCO Settings.

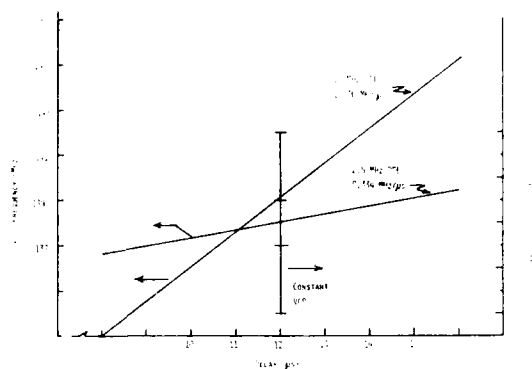


Fig. 5 Plot of the Output Delay vs the VCO Frequency and vs the Input Frequency for Two Different Bandwidth Filters.

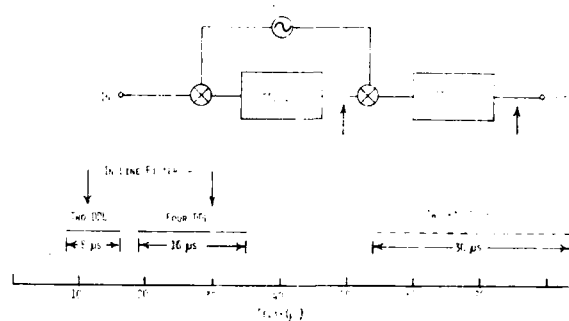


Fig. 6 Schematic of the Variable Delay Line (Top) and Plot of the Variable Delay Range (Bottom) using "In-Line" and RAC Filters.

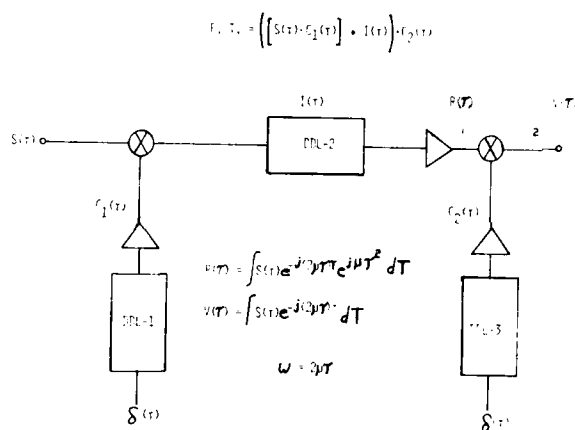


Fig. 7 Schematic of the Chirp Transform Processor.

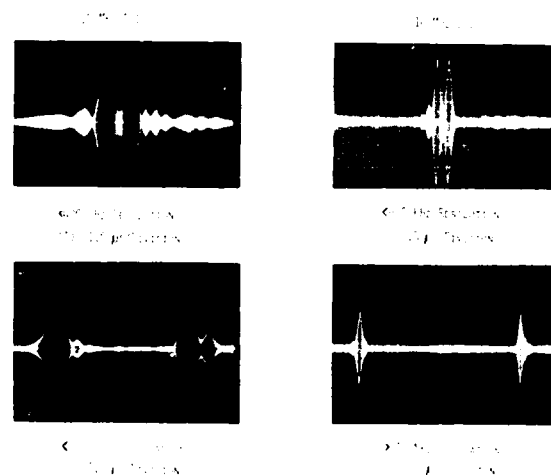


Fig. 8 Photograph of Resolution and Maximum Frequency Separation of the Chirp Transform Processor Using Two Different Bandwidth Filters.

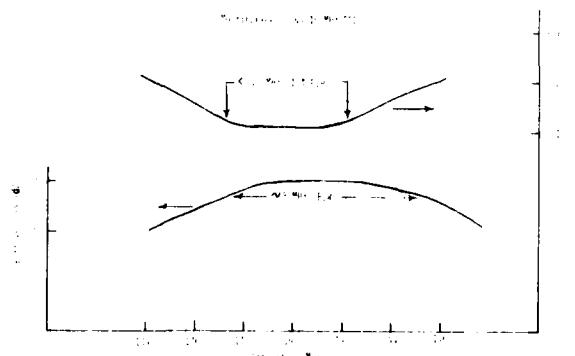


Fig. 9 Plot of the Amplitude Bandwidth and Compressed Pulse Bandwidth of the "Time-Ordered" Multiplexer Using 10 MHz Chirp Filters.

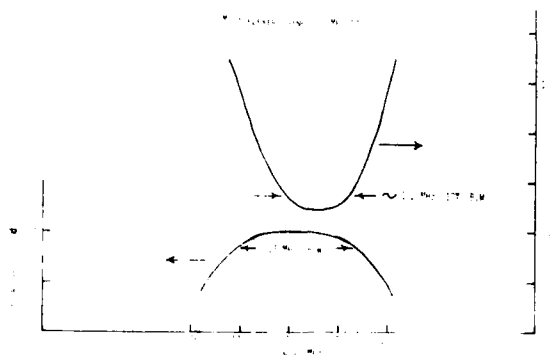


Fig. 10 Plot of the Amplitude Bandwidth and Compressed Pulse Bandwidth of the "Time-Ordered" Multiplexer Using 2.5 MHz Chirp Filters.

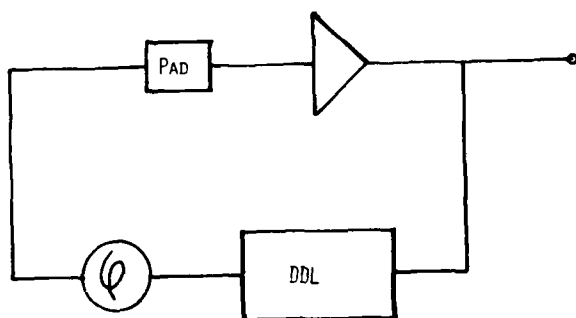


Fig. 11 Schematic of the Chirp Oscillator.

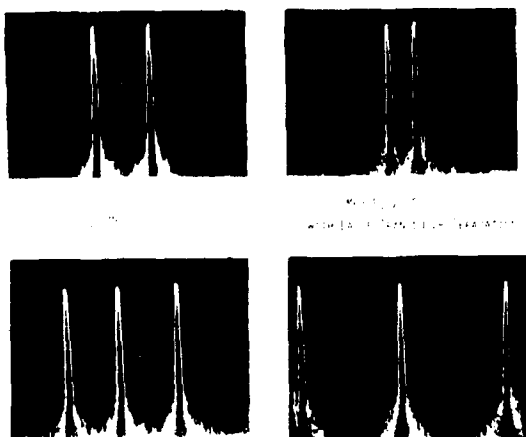
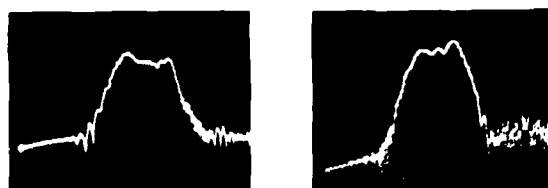


Fig. 13 Spectrum Analyzer Photographs Showing the Tuning Range of Two Chirp and Two Delay Line Oscillators for Approximately a 360 Degree Phase Shift.



(A) PASS - BAND OF CONJUGATE FILTER

(B) PASS - BAND OF EXPANDER FILTER

$$H = \frac{1}{\sqrt{2}} \left(\frac{1}{\sqrt{2}} \right)$$

$$L = \frac{1}{\sqrt{2}} \left(\frac{1}{\sqrt{2}} \right)$$

$$Q = \frac{1}{\sqrt{2}}$$

Fig. 14 Photographs of the Passband Response of (A) Conjugate 2.5 MHz Chirp Filter, and (B) Expander 2.5 MHz Chirp Filter.

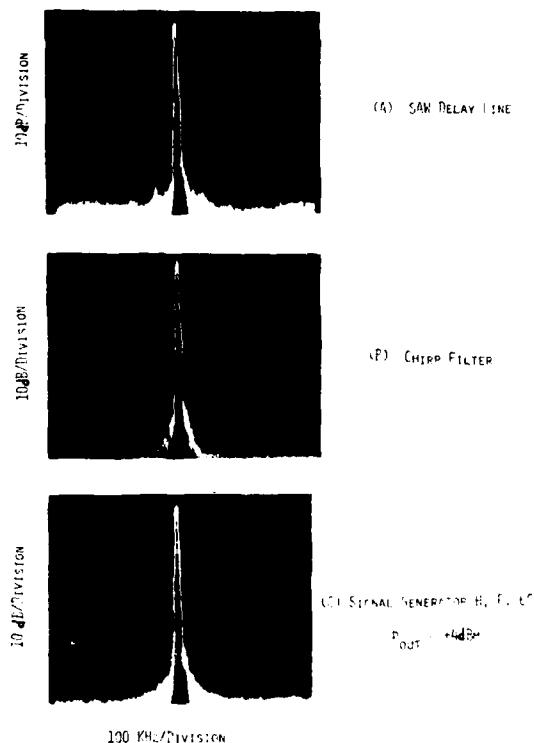


Fig. 12 Spectrum Analyzer Photographs of (A) SAW Delay Line Oscillator, (B) Chirp Filter Oscillator, and (C) Signal Generator Set at a Fixed Frequency.

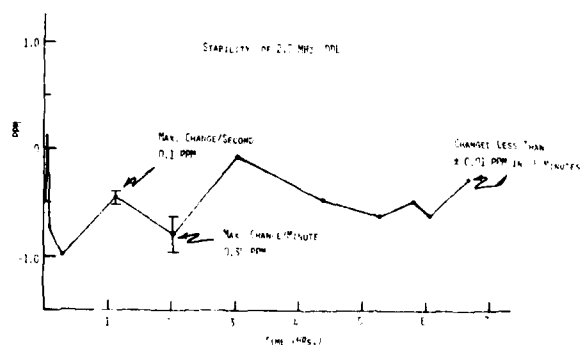


Fig. 15 Plot of Stability in PPM of the Conjugate 2.5 MHz Chirp Filter Oscillator.

OSCILLATOR SPECIFICATIONS : A REVIEW OF CLASSICAL AND NEW IDEAS

Jacques Rutman

Laboratoire Primaire du Temps et des Fréquences (LPTF)

Bureau National de Métrologie et Observatoire de Paris
75014 Paris, France

Summary

This paper will attempt at giving an insight into the various parameters that have been proposed as measures of frequency stability. First, well-documented classical parameters which have gained wide acceptance will be recalled. Emphasis will be put on the transfer function that links any statistical time-domain parameter to the spectral density of the frequency fluctuations.

Then, other parameters which have been independently proposed as possible substitutes for classical time domain parameters by several authors will be presented, using a common terminology throughout the paper. The review will cover the "transfer function approach" (Rutman), the method of the "finite time frequency control" (Picinbono et al), the time domain parameter $E[\delta f_m^2]$ (De Prins et al) and the "structure function approach" (Lindsey et al). The transfer function will be given for each parameter in order to show their respective properties. The links between the various approaches, and also the links with classical parameters will be shown.

In particular, a three-sample modified sample variance will be defined which has specific advantages relatively to the classical two-sample variance ; it will be linked to the 3rd structure function of phase noise. Also, the time domain parameter $E[\delta f_m^2]$ will be linked to the first structure function of frequency noise and to the two-sample variance with non adjacent samples. The Nth structure function of phase noise will be shown to be nothing but the Hadamard variance weighted by binomial coefficients (Baugh).

Introduction

Since many decades, the ever-increasing use of electromagnetic waves has created a need for better and better frequency control in frequency sources. Very high stability sources have been developed, such as quartz crystal oscillators, atomic frequency standards and more recently saturated-absorption frequency stabilized lasers. These devices are now used in a very great number of scientific and technical applications. Hence, many people (oscillator manufacturers, scientists, users, etc.) need very precisely defined specifications for frequency stability in order to make meaningful comparisons among various sources, and also to be able to predict the performance of systems using these sources. Many papers have been devoted to the characterization of frequency (in)stability during the last fifteen years. Before presenting the goals of this paper which deals mainly with instability due to random processes, let us recall the histo-

ry of frequency stability characterization where four main periods may be considered :

Before 1964, the need for good short-term frequency stability arose in the development of many systems such as Doppler radar, coherent radar, missile and spacecraft tracking and guidance systems, communication systems, frequency standards, etc. Much work was done independently by many groups but communication between them was severely hampered by the lack of common ideas, concepts and terminology. At the beginning of the sixties, the need for clarification was obvious.

1964-1966. The first (and unique up to now) symposium devoted solely to the subject of short term frequency stability was held in November 1964¹. Following this symposium, a technical subcommittee dealing with frequency stability has been created within the IEEE, the ultimate aim being an IEEE standard on the definition and measurement of both short-term and long-term stability. A special issue of the Proceedings of the IEEE devoted to frequency stability was published in February 1966². It is during that period that the first basic papers dealing explicitly with concepts such as time domain and frequency domain characterization were published^{3, 4, 5}.

1966-1971. After 1966, many people have used some of the parameters introduced in references 1 and 2. But some important questions still needed a clear answer, such as : what measurement process should be used to get an estimate of the variance of frequency fluctuations averaged over a time interval τ ? In 1970-1971, the IEEE Subcommittee gave some answers by recommending two measures for frequency stability, $S_y(f)$ in the Fourier frequency domain and $\sigma_y(\tau)$ in the time domain^{6, 7}. During that period, many scientists around the world realized that much work (both theoretical and experimental) was still necessary in order to get a good understanding of the problems involved in oscillator frequency stability characterization and measurement.

After 1971, the two recommended parameters have gained wide acceptance in the time and frequency community⁸, especially the square root of the two-sample zero-dead-time (Allan) variance $\sigma_y^2(\tau)$. But during the same period, several authors have independently proposed other approaches with new parameters which are believed to exhibit some specific advantages when compared to a classical parameter such as $\sigma_y^2(\tau)$. Also, some authors have pointed out that mathematical mistakes had been made in the pioneering papers of 1964-1966, although many useful practical results were obtained.

The goals of this paper are :

- to recall briefly some of the mathematical problems that arise when dealing with random models for phase and frequency noises in oscillators.
- after a description of the classical parameters, to present the non-classical concepts and parameters that have been proposed by several authors^{10, 22, 23, 24}, including the relationships between them when they exist, and with $S_V(f)$ and/or $\sigma_V(\tau)$. In particular, emphasis will be put on the transfer function that allows one to link any time domain parameter to the spectral density $S_V(f)$.

With new sources to be developed and new applications of ultra-stable oscillators, a need for more sophisticated parameters than the classical ones used to-day may arise in the future. Also, these new parameters may be better tools for studying the statistical properties of noise in oscillators.

Mathematical models for phase and frequency noises.

At first, it is important to emphasize on the fact that the two following facets are often confused^{9, 32} :

- the real world, with its physical devices, measurement apparatus, experimental results derived from reading meters, counters, dial settings and so on.
- the mathematical model, with the means and rules for operating with the symbols introduced in it.

In an ideal world, the correspondence between the physical reality and the model should be perfect. In fact, the real world is so complex that many details are ignored in the model and therefore, the correspondence is incomplete and imprecise. On the other hand, some properties that have no direct meaningful counterparts in the real world have to be introduced in the model in order to make a tractable model (eg : stationarity properties of random processes). In any case, the model must be well defined from the mathematical point of view.

For oscillator specifications, the model which has been found useful in order to represent the instantaneous quasi-sinusoidal output signal reads as^{7, 32} :

$$V(t) = [V_0 + \epsilon(t)] \sin[2\pi\nu_0 t + \varphi(t)] \quad (1)$$

V_0 and ν_0 are the nominal amplitude and frequency respectively ; $\epsilon(t)$ and $\varphi(t)$ are random processes denoting amplitude noise and phase noise respectively. Amplitude noise does not contribute directly to phase noise, although AM to PM noise conversion can occur in some non-linear devices. Very often, $\epsilon(t)$ is neglected and the model reads as :

$$V(t) = V_0 \sin[2\pi\nu_0 t + \varphi(t)] \quad (2)$$

The key element in this model is the random process $\varphi(t)$ whose statistical properties have to be specified (eg stationarity or non stationarity) before doing calculations or using results from the random process theory. In working with this model of $V(t)$, it is com-

mon to introduce the instantaneous angular frequency of the signal defined as the time derivative of the total instantaneous phase^{3, 4, 5, 7} :

$$\omega(t) = \frac{d}{dt} (\omega_0 t + \varphi(t)) = \omega_0 + \frac{d\varphi(t)}{dt} \quad (3)$$

where $\dot{\varphi}(t) = \frac{d\varphi(t)}{dt}$ is a random process denoting

angular frequency fluctuations around the value ω_0 of the ideal noiseless oscillator. Very often in the past, both $\varphi(t)$ and $\dot{\varphi}(t)$ have been assumed to be simultaneously zero mean, stationary random processes. In fact, things are not so simple when dealing with derivative or integral of random processes^{10, 11}. Let us consider some of the problems that may appear.

Stationarity of $\varphi(t)$.

As far as the model is concerned, it is very convenient to assume a priori the stationarity (at least at the second order) of $\varphi(t)$, since much of the theory or random processes is only valid in this case, especially spectral analysis using the correlation function and the spectral density¹¹. But the theoretical analysis of internal white noise in oscillators leads to a phase diffusion process analog to the Brownian motion which is clearly a non-stationary process^{12, 13, 14} ; in this case, it is impossible to introduce rigorously a correlation function or a spectral density of the phase, although it has been made by many authors. More generally, when physical arguments show that it is reasonable to use a stationary model for $\dot{\varphi}(t)$, it can be seen from the integral relationships between $\dot{\varphi}(t)$ and $\varphi(t)$ that the latter is not always a stationary process with integrable spectral density¹⁰. When dealing with the power law spectral density model^{5, 7} which has been found very useful to represent the various types of noise in sources, one has to be aware of the problems due to the non-integrability of these laws and hence of their lack of significance as spectral densities when they are extrapolated down to the zero frequency.

Existence of $\dot{\varphi}(t)$.

Another problem when dealing with models for phase and frequency noises is that the quantity $\dot{\varphi}(t)$, whose physical interpretation seems obvious, is not always mathematically defined since many random processes have no derivative at least in the function sense (eg an oscillator with phase jumps at random instants, or with a phase diffusion process)¹⁰.

As a rule, a stationary process $\varphi(t)$ is differentiable in the mean square sense if its correlation function $R_\varphi(\tau)$ has derivatives of order up to two ; the correlation function of $\dot{\varphi}(t)$ reads then as¹¹ :

$$R_{\dot{\varphi}}(\tau) = - \frac{d^2 R_\varphi(\tau)}{d\tau^2} \quad (4)$$

In conclusion, one has to be very careful when dealing with mathematical models for phase and frequency noises since one may be led to use non-existing quantities such as $\dot{\varphi}(t)$ or $S_{\dot{\varphi}}(f)$. The reader is referred to the references for more details.

Characterization of frequency stability : a review of classical ideas.

Sources manufacturers and users are obviously interested in the stability of the output frequency and hence, as far as the model is concerned, with the statistical properties of $\dot{y}(t)$, assuming that this process exists. We now survey the classical parameters that have been introduced as measures of frequency stability^{3, 4, 5, 7}. The Hadamard variance¹⁵ which is used by several laboratories will be also discussed here.

Fourier frequency domain.

Any stationary random process may be characterized in the Fourier frequency domain by its spectral density which is the Fourier transform of its correlation function¹¹. Let us denote the frequency fluctuations by $\Delta y(t) = \frac{\dot{y}(t)}{2\pi}$ and the fractional instantaneous (angular) frequency noise by :

$$y(t) = \frac{\dot{y}(t)}{\omega_0} = \frac{\Delta y(t)}{\nu_0} \quad (5)$$

Assuming that $\dot{y}(t)$ exists and that it is stationary, it is possible to consider the following spectral densities⁷ $S_{\dot{y}}(f)$, $S_{\Delta y}(f)$ and $S_y(f)$. If $y(t)$ is also a stationary process, its spectral density $S_y(f)$ is defined unambiguously. The relations between these spectral densities read as :

$$\begin{aligned} S_{\dot{y}}(f) &= 4\pi^2 f^2 S_y(f) = 4\pi^2 S_{\Delta y}(f) \\ S_y(f) &= \frac{1}{\nu_0^2} S_{\Delta y}(f) = \left(\frac{f}{\nu_0}\right)^2 S_{\dot{y}}(f) \end{aligned} \quad (6)$$

The IEEE subcommittee has recommended the use of the one-sided spectral density $S_y(f)$ on a per Hertz basis as a measure for frequency stability in the Fourier frequency domain^{6, 7}. The actual experimental estimations of spectra can be quite elaborate.

Time domain.

Although the spectral density $S_y(f)$ contains the whole information on the relative frequency stability, experimenters often need a quantitative measure of the level of frequency fluctuations measured over a time interval τ : this is the primary goal of a time domain parameter. Since random phenomena are involved, many measurements have to be made, each of duration τ , before calculating a statistical meaningful parameter such as the variance.

In the following, we will use mainly $y(t)$ which is assumed to be a zero mean stationary random process. Any experimental apparatus devised to measure the (fractional) instantaneous frequency always involves a finite averaging time interval τ ; the result of one measurement of duration τ beginning at t_k will be denoted as :

$$\bar{y}_k = \frac{1}{\tau} \int_{t_k}^{t_k+\tau} y(\theta) d\theta \quad (7)$$

Obviously, \bar{y}_k is a random variable and statistical parameters have to be used to measure the dispersion of \bar{y}_k when many measurements are made.

We present here the various parameters that have

been suggested as measures of frequency stability in the time domain. Several cases have to be considered according to whether the data are continuous or discrete, infinite or finite in length.

Continuous data and infinite data length.

One of the first parameters that has been suggested^{3, 4} as a measure for frequency stability over a time interval τ is the true variance of \bar{y}_k defined as* :

$$I^2(\tau) = \sigma^2[\bar{y}_k] = \langle \bar{y}_k^2 \rangle \quad (8)$$

This definition implies continuous data and infinite data length; more practical situations will be considered later.

The relation between $I^2(\tau)$ and $S_y(f)$ reads as^{3, 4}:

$$I^2(\tau) = \int_0^\infty S_y(f) \left(\frac{\sin \pi \tau f}{\pi \tau f} \right)^2 df \quad (9)$$

This relation shows the limitations of the true variance $I^2(\tau)$ as a practical measure for frequency stability: since the transfer function that appears in the integral behaves as 1 for $\pi \tau f \ll 1$, $I^2(\tau)$ is very sensitive to the low Fourier frequency components of $S_y(f)$ which may be due to more or less random slow drifts. Moreover, the integral diverges for the flicker frequency noise model ($S_y(f) \sim f^{-1}$), whereas flicker frequency noise has been found to exist in all sources, at least over a finite Fourier frequency range going down to very low frequencies. That is why the true variance is an idealization whose utility is very limited for practical characterization of real oscillators. One way to get rid of these problems is to consider the more realistic case of finite data length⁴.

Continuous data and finite data length.

The definition of $I^2(\tau)$ assumes that $y(t)$ exists for all time but in practice, real signals only exist over some finite time interval T_0 . Under these conditions, one can only make estimates of true statistical parameters, these estimates being themselves random variables with average values, variances and so on. In our problem, the average value $I^2(\tau; T_0)$ of the estimate of $I^2(\tau)$ made over a time interval T_0 is related to $S_y(f)$ by⁴:

$$I^2(\tau; T_0) = \int_0^\infty S_y(f) \left(\frac{\sin \pi \tau f}{\pi \tau f} \right)^2 \left(1 - \frac{\sin^2 \pi (T_0 - \tau) f}{(\pi (T_0 - \tau) f)^2} \right) df \quad (10)$$

It differs from equation (9) in that the low frequency part of the spectral density (below about T_0^{-1}) is effectively removed before the integration^{**}. In particular, the integral converges for the flicker frequency noise model since the transfer function in (10) behaves as $\frac{\pi^2 (T_0 - \tau)^2}{3} f^2$ for $\pi (T_0 - \tau) f \ll 1$.

Let us now consider the case of discrete data which is closer to the physical reality when measurements are made with digital counters.

* As only one (or a few) oscillator is usually available to the experimenter for measurement, one will use time average of random quantities, the infinite time average of x being denoted as $\langle x \rangle$.

** $T_0 - \tau$ appears in place of T_0 in (10), since only $T_0 - \tau$ is available for averaging.

Discrete data.

As the period or frequency counting technique using digital counters is very often used for data acquisition, estimates of $I^2(\tau)$ or $I^2(\tau; T_0)$ have to be made from a finite number of discrete samples of \bar{y}_k . An analysis of this estimation process involving the sample variance has been developed by Allan⁵ and will be recalled here. Let us consider a measurement cycle involving N discrete measures of \bar{y}_i , with τ the averaging time, T the time interval between the beginnings of two successive measurements ($T = \tau + \text{dead time between successive measurements}$), and f_M the measurement bandwidth (figure 1). The sample variance is defined as* :

$$\sigma_y^2(N, T, \tau, f_M) = \frac{1}{N-1} \sum_{i=1}^N \left(\bar{y}_i - \frac{1}{N} \sum_{i=1}^N \bar{y}_i \right)^2 \quad (11)$$

It is itself a random variable whose infinite time average $\langle \sigma_y^2(N, T, \tau, f_M) \rangle$ has to be considered in order to get a statistically significant measure of frequency stability.

For noises for which $I^2(\tau)$ exists, it is the limit of $\langle \sigma_y^2(N, T, \tau, f_M) \rangle$ as $N \rightarrow \infty$. But in many cases, it would be wrong to assume that $\langle \sigma_y^2(N, T, \tau, f_M) \rangle$ converges to a meaningful limit as $N \rightarrow \infty$ ^{5, 7, 18}. The following relation has been established by Cutler⁷ :

$$\langle \sigma_y^2(N, T, \tau) \rangle = \frac{N}{N-1} \int_0^{\infty} S_y(f) \left(\frac{\sin \pi \tau f}{\pi \tau f} \right)^2 \left(1 - \frac{\sin^2 N \pi T f}{N^2 \sin^2 \pi T f} \right) df \quad (12)$$

For $N \pi T f \ll 1$, the transfer function behaves as $\frac{N(N+1)}{3} (\pi T f)^2$ and the integral converges for the flicker frequency noise model.

The two-sample (Allan) variance.

In order to improve comparability of data, it is important to specify N , T , τ and f_M with any results. In 1971, the IEEE subcommittee has recommended the choice of the two-sample ($N = 2$), zero-dead-time ($T = \tau$) sample variance^{6, 7}. The proposed measure of frequency stability in the time domain reads then as** :

$$\sigma_y^2(\tau) = \langle \sigma_y^2(2, \tau, \tau, f_M) \rangle = \frac{\langle (\bar{y}_{k+1} - \bar{y}_k)^2 \rangle}{2} \quad (13)$$

$\sigma_y^2(\tau)$ is an idealization in the sense that only estimates of the infinite time average can be obtained experimentally over a finite total duration time¹⁶. But even

* The factor $\frac{1}{N-1}$ was chosen instead of $\frac{1}{N}$ in order to get an unbiased estimate of the true variance for white frequency noise, for any finite value of $N \geq 2$. Other factors may be chosen in order to suppress the bias for other noises¹⁸. Often, f_M is not explicitly stated in the formulas.

** No recommendation was made for the value of f_M to be used. As $\sigma_y^2(\tau)$ may depend on f_M in some cases, it is important to specify f_M with any experimental result (1 or 2 kHz BW is often used).

as an idealization, it has greater utility than $I^2(\tau)$ since it converges for noise processes that do not have convergent $\langle \sigma_y^2(N, T, \tau, f_M) \rangle$ as $N \rightarrow \infty$ (particularly for flicker frequency noise).

For $N = 2$ and $T = \tau$, equation (12) reduces to¹⁷ :

$$\sigma_y^2(\tau) = \int_0^{\infty} S_y(f) \frac{2 \sin^4 \pi \tau f}{(\pi \tau f)^2} df \quad (14)$$

where the transfer function behaves as $2(\pi \tau f)^2$ for $\pi \tau f \ll 1$, explaining thus the convergence for $S_y(f) \sim f^{-1}$ (see figure 2).

The plots giving $\sigma_y(\tau)$ versus τ are now very familiar to everyone working in the field. In certain cases, the laws of $\sigma_y(\tau)$ versus τ enables one to deduce the kinds of noise that exist in the source⁷, but there are some limitations due to the fact that the main lobe of the transfer function shown on figure 2 is too wide to allow high resolution spectral analysis of $y(t)$. As an example, both white and flicker phase noises ($S_y(f) = \text{constant}$ and $\sim f^{-1}$, respectively) lead to a τ^{-1} law for $\sigma_y(\tau)$. In fact, the two-sample variance is a useful time domain parameter whose primary use is not spectral analysis (although it is possible for some power laws).

The Hadamard variance.

Since the spectral density is a very important parameter, and since time domain measurements of \bar{y}_i with counters are very convenient, Baugh has introduced a statistical parameter, the Hadamard variance¹⁵, which allows one to make estimates of $S_y(f)$ from measurements of \bar{y}_i . The Hadamard variance is defined from ensembles of N measurements of \bar{y}_i as follows (see figure 3a) :

$$\langle \sigma_H^2(N, T, \tau) \rangle = \langle (\bar{y}_1 - \bar{y}_2 + \bar{y}_3 - \dots - \bar{y}_N)^2 \rangle \quad (15)$$

As any other time domain parameter, it may be written under the general form :

$$\langle \sigma_H^2(N, T, \tau) \rangle = \int_0^{\infty} S_y(f) |H_N(f)|^2 df \quad (16)$$

where $|H_N(f)|^2$ is now given by¹⁹ (see figure 3b)

$$|H_N(f)|^2 = \left(\frac{\sin \pi \tau f}{\pi \tau f} \right)^2 \left(\frac{\sin N \pi T f}{\cos \pi T f} \right)^2 \quad (17)$$

Although this transfer function has a narrow main lobe centered at Fourier frequency $f_1 = \frac{1}{2T}$, it has two

limitations :

. when $T = \tau$ (adjacent samples), spurious responses exist at odd harmonics of f_1 ; this can be partly eliminated with an optimum dead-time between samples¹⁵.

. the transfer function has large sidelobes around the main lobe; these can be modified or even eliminated by multiplying each \bar{y}_i within the set of N measurements by an appropriate weighting coefficient; the binomial coefficients completely eliminates the side lobes¹⁵. The Hadamard variance weighted by binomial

coefficients (B. C.) reads as*:

$$\langle \sigma_{H_{bc}}^2(N, T, \tau) \rangle = \left\langle \left(\sum_{i=1}^N (-1)^{i-1} \binom{N-1}{i-1} \bar{y}_i \right)^2 \right\rangle \quad (18)$$

with N even.

In this case, the transfer function is given by^{19, 20}:

$$|H_{bc}(f)|^2 = 2^{N-1} \left(\frac{\sin \pi \tau f}{\pi \tau f} \right)^2 \sin^{N-1} \pi \tau f \quad (19)$$

Figure 4 shows this function for $T = \tau$ and $N = 10$.

Experimentally, the Hadamard variance is very useful since it extends frequency domain measurements down to very low Fourier frequencies using a digital counter working in the time domain with long τ . Other improvements have been proposed in order to suppress some of the spurious harmonic responses^{19, 20, 21}.

New ideas for characterization of frequency stability

Following the recommendation of the IEEE subcommittee, the two-sample variance has been very widely used as a time domain measure of frequency stability. However, several authors have found useful to develop other approaches leading to new parameters which are believed to be more "efficient" than $\sigma^2(\tau)$ in some respect. The Hadamard variance was a first example of a parameter with greater spectral analysis capability.

In this section, we present the main aspects** of these other approaches while trying to establish the links between them, and with the classical parameters. Emphasis will be put on the transfer function $H(f)$ since each new parameter $\sigma^2(\tau)$ may be written as:

$$\sigma^2(\tau) = \int S_y(f) |H(f)|^2 df \quad (20)$$

Instead of using the various symbols used in the original papers, we will use here the terminology usually used in the time and frequency literature⁷ except when new symbols are necessary for clarity. The review will cover approaches proposed by Rutman²², Picinbono et al¹⁰, De Prins et al²³ and Lindsey et al²⁵.

The transfer function approach²²

Classically, any time domain parameter $\sigma^2(\tau)$ is defined by its measurement sequence in the time domain (involving quantities such as N , T , τ and possibly weighting factors). Then the relationship between $\sigma^2(\tau)$ and $S_y(f)$ is established and reads as equation (20), where the transfer function $H(f)$ is the Fourier transform of a discontinuous function of time which resembles the measurement sequence (as an example, see figures 2 and 3). Hence, the transfer function results unambiguously from the measurement cycle.

In the transfer function approach, a time domain parameter is defined by its transfer function even if no practical measurement sequence exists in the time domain (i.e. implementable with a digital counter). The theoretical interest of this approach is that it leads to a better understanding of the rela-

tionships between time domain and frequency domain measures²². The practical interest is that it leads to an easily implementable experimental test set using the phase detector technique: the defined parameter is measured by analog filtering the demodulated phase noise at the output of a mixer driven with two signals at the same nominal frequency and placed in quadrature²⁶.

Two new variances have been defined using this approach.

The high-pass variance $\sigma_{HP}^2(\tau)$

It is defined by high-pass filtering the demodulated phase noise $\varphi(t)$ with a filter whose low frequency cutoff is $f_c = (\pi\tau)^{-1}$. More precisely***:

$$\sigma_{HP}^2(\tau) = \frac{8}{\omega_0^2 \tau^2} \int_0^\infty S_y(f) |H_{HP}(f)|^2 df \quad (21)$$

The detailed study of $\sigma_{HP}^2(\tau)$ has shown that it has the same general behavior as $\sigma^2(\tau)$, except for a numerical constant of the order of unity. Particularly, both have the same laws versus τ and the same dependence (or independence) versus f_M for the power law spectral density model. Also, the high pass behavior of $\sin^4 \pi \tau f$ explains clearly why $\sigma^2(\tau)$ is not selective enough to give two different laws versus τ for white and flicker phase noises for which $\sigma_y(\tau) \sim \tau^{-1}$. This parameter shows clearly that the stability over a time interval τ is equally due to all Fourier frequency components of $S_y(f)$ lying above $f_c = (\pi\tau)^{-1}$, and hence that the periodic behavior of $\sin^4 \pi \tau f$ in equation

$$\sigma_y^2(\tau) = \frac{8}{\omega_0^2 \tau^2} \int_0^\infty S_y(f) \sin^4 \pi \tau f df \quad (22)$$

is not critically important for the general behavior of $\sigma^2(\tau)$. A bandpass filtering of the demodulated frequency noise is also possible³⁴ to estimate $\sigma_y(\tau)$.

The bandpass variance.

Having recognized the origin of the lack of selectivity of $\sigma_y^2(\tau)$ or $\sigma_{HP}^2(\tau)$, it is interesting to define a new variance in such a way that its transfer function be selective, i.e. bandpass. The first attempt through a more complicated measurement sequence in the time domain was the Hadamard variance with its inherent limitations. The transfer function approach allows one to define a bandpass variance in a logical way as:

$$\sigma_{BP}^2(\tau) = \frac{8}{\omega_0^2 \tau^2} \int_0^\infty S_y(f) |H_{BP}(f)|^2 df \quad (23)$$

where $H_{BP}(f)$ is the transfer function of a bandpass

* The weighting by binomial coefficients appears in a logical way by considering the N th phase increment introduced in 1966 by Barnes³³ and used again in 20, 24, 25.

** For more details, the reader is referred to the original papers.

*** By analogy with the relation between $\sigma_y^2(\tau)$ and $S_y(f)$, see equation (22).

filter whose center frequency is equal to $f_0 = (2\tau)^{-1}$ and with constant Q factor as f_0 varies.

The selectivity is now sufficient to distinguish between white phase noise and flicker phase noise for which $\sigma_{\phi}^2(\tau) \sim \tau^{-3}$ and $\sigma_{\phi}^2(\tau) \sim \tau^{-2}$ respectively (with no dependence on f_H). For other noises such as white or flicker frequency noises, $\sigma_{\phi}^2(\tau)$ is an estimate of the two-sample variance.

Obviously, the bandpass variance is nothing but an analog constant percentage bandwidth spectral analysis of phase noise presented in the time domain through equation (23). In a certain sense, the bandpass variance is the ideal ultimate goal which is pursued by improving the Hadamard variance^{19, 20, 21} since all the improvements are made in order to get a narrow bandpass transfer function without any side lobes or spurious harmonic responses.

Method of the Finite-Time Frequency Control¹⁰

Continuous data.

Noting the limitations of $\sigma_y^2(\tau)$ with regard to the possibility of separating "slow" fluctuations from "fast" ones (i. e. to speak of "short" term and "long" term stability) and with regard to the possibility of making an accurate spectral analysis of $y(t)$, Picinbono et al¹⁰ have developed a method based on the study of the quantity :

$$\bar{Y}_{t, T_0} = \bar{y}_t - \frac{1}{T_0} \int_{t-\frac{T_0}{2}}^{t+\frac{T_0}{2}} \bar{y}_t dt \quad (24)$$

which is obtained by subtracting from \bar{y}_t its time average over an observation time T_0 around t (not to be confused with T appearing in $\sigma^2(N, T, \tau)$). The variance of this quantity is a modified version of the true variance $I^2(\tau)$ given by equation (8), which is now also a function of T_0 ; let us denote it as :

$$J^2(\tau; T_0) = \sigma^2[\bar{Y}_{t, T_0}] \quad (25)$$

Since $y(t)$ is assumed to be a zero-mean process, the true variance is the limit of $J^2(\tau; T_0)$ when $T_0 \rightarrow \infty$.

The interest of $J^2(\tau; T_0)$ with respect to other parameters such as $I^2(\tau)$, $I^2(\tau; T_0)$ or $\sigma^2(\tau)$ comes from its relationship to $S_y(f)$, since¹⁰ :

$$J^2(\tau; T_0) = \int_0^\infty S_y(f) \left(\frac{\sin \pi \tau f}{\pi \tau f} \right)^2 \left(1 - \frac{\sin \pi T_0 f}{\pi T_0 f} \right)^2 df \quad (26)$$

This expression differs from equation (9) in that the low frequency portion of the spectrum is removed before the integration. For $\pi T_0 f \ll 1$, the transfer function in (26) behaves as $\frac{1}{\pi T_0 f}$ and the filtering is hence more abrupt than the one appearing in equation (10) giving $I^2(\tau; T_0)$.

More precisely, for the power law spectral model

$S_y(f) = K_\alpha / f^\alpha$, $I^2(\tau)$ is finite for α up to $\alpha < 1$, $I^2(\tau; T_0)$ and $\langle \sigma_y^2(N, T, \tau) \rangle$ for α up to $\alpha < 3$, whereas $J^2(\tau; T_0)$ is finite for α up to $\alpha < 5$. Hence, the finite-time frequency control method allows one to study spectral densities with greater exponent ("stronger nonstationarity") without divergence problems (at least for $\alpha < 5$). However, at the

best of my knowledge, frequency noise processes with $\alpha > 2$ have not been reported for real sources over the Fourier frequency range covered by existing experimental measurement systems. Besides, $J^2(\tau; T_0)$ is believed to be more adequate than $\sigma_y^2(\tau)$ for separating "slow" from "fast" fluctuations in $y(t)$, and also for the precision obtained by finite-time measurements¹⁰.

Discrete data.

Since digital counters give only discrete samples \bar{y}_i , an estimate of $J^2(\tau; T_0)$ using N values of \bar{y}_i ($i = 1$ to N) must be considered. With N odd, a modified sample variance may be written as :

$$\Sigma_y^2(N, T, \tau) = \left(\bar{y}_{\frac{N+1}{2}} - \frac{1}{N} \sum_{i=1}^N \bar{y}_i \right)^2 \quad (27)$$

N being odd, $\bar{y}_{\frac{N+1}{2}}$ denotes the central sample within the set of N samples. $\Sigma_y^2(N, T, \tau)$ is itself a random variable the infinite time average of which is related to $S_y(f)$ by¹⁰ :

$$\langle \Sigma_y^2(N, T, \tau) \rangle = \int_0^\infty S_y(f) \left(\frac{\sin \pi \tau f}{\pi \tau f} \right)^2 \left(1 - \frac{\sin \pi N T f}{N \sin \pi T f} \right)^2 df \quad (28)$$

For $N T f \ll 1$, the transfer function behaves as

$$\frac{(N^2 - 1)^2 (\pi T f)^4}{36} \quad \text{and the convergence exists for } \alpha < 5$$

A modified three-sample variance.

With arguments similar to those used for the classical sample variance $\sigma_y^2(N, T, \tau)$, it is interesting to consider a particular modified sample variance $\Sigma_y^2(N, T, \tau)$, namely the one with $T = \tau$ (adjacent samples) and $N = 3$ (the smallest possible value of N with this parameter). Then, equations (27) and (28) reduces respectively to :

$$\Sigma_y^2(\tau) = \langle \Sigma_y^2(3, \tau, \tau) \rangle = \frac{1}{9} \langle (2\bar{y}_2 - \bar{y}_1 - \bar{y}_3)^2 \rangle \quad (29)$$

$$\Sigma_y^2(\tau) = \int_0^\infty S_y(f) \frac{16}{9} \frac{\sin^2 \pi \tau f}{(\pi \tau f)^2} df \quad (30)$$

The transfer function is such that $\Sigma_y^2(\tau)$ is convergent for α up to $\alpha < 5$. It is easy to see that

$\frac{16}{9} \frac{\sin^2 \pi \tau f}{(\pi \tau f)^2}$ is the square of the Fourier transform of the measurement sequence of

$\frac{1}{3}(2\bar{y}_2 - \bar{y}_1 - \bar{y}_3)$ (see figure 4). For values of α where $\sigma_y^2(\tau)$ converges, $\Sigma_y^2(\tau)$ and $\sigma_y^2(\tau)$ have the same laws versus τ and f_H and very close numerical values.

Time domain approach using $E[\delta y_m^2]$ ²³.

Another time domain approach has been suggested by De Prins et al which is a measure of the dispersion of instantaneous frequency measurements made over an ensemble of sources at time $t_0 + mT_s$ (m integer), relatively to the reference time t_0 .

*It is the center frequency of the first lobe of $\sin^2 \pi \tau f$ for which $Q = 1, 37$.

Let us assume that $y(t)$ is sampled at $t_0, t_1 = t_0 + T_s, \dots, t_m = t_0 + mT_s$, with a zero averaging time τ . This sampling gives a set of instantaneous values of $y(t)$: y_0, y_1, \dots, y_m . The parameter is then defined as *

$$E[\delta y_m^2] = E[(y_m - y_0)^2] \quad (31)$$

Since only one sample of $y(t)$ is usually available to the experimenter, this parameter has to be estimated according to the measurement sequence described in²⁷.

The relation between this parameter and $S_y(f)$ reads as²⁷:

$$E[\delta y_m^2] = \int_0^\infty S_y(f) 4 \sin^2 m\pi T_s f df \quad (32)$$

where the transfer function is now a periodic undamped waveform. Application of this relation with the power law spectral density model allows one to study the evolution of $E[\delta y_m^2]$ with m , which represents the evolution of the dispersion of results as time passes^{23, 27}. This is quite different from the evolution of a variance with averaging time τ (which is assumed to be zero in (31)). In particular, for white frequency noise, $E[\delta y_m^2]$ is constant, whereas for flicker frequency noise, a logarithmic variation in time appears²⁷.

Relation with the sample variance.

Although the approaches are quite different from the conceptual point of view, a simple relation may easily be written between $E[\delta y_m^2]$ and the sample variance since for $N = 2$ and $\tau = 0$, equation (12) may be rewritten as:

$$\langle \sigma_y^2(2, T, 0) \rangle = 2 \int_0^\infty S_y(f) \sin^2 \pi T f df \quad (33)$$

And hence, denoting $T = mT_s$, one gets **:

$$E[\delta y_m^2] = 2 \langle \sigma_y^2(2, T, 0) \rangle \quad (34)$$

Non-zero averaging time τ .

By keeping the same philosophy as above, it is possible to introduce a non-zero averaging time τ in the definition of the parameter, being thus closer to practical measurement sequences. Then, y_0 and y_m in (31) are replaced by \bar{y}_0 and \bar{y}_m (see equation (7)) and the new parameter, denoted $I_y^2(T, \tau)$, is equal to twice $\langle \sigma_y^2(2, T, \tau) \rangle$:

$$I_y^2(T, \tau) = 4 \int_0^\infty S_y(f) \left(\frac{\sin \pi \tau f}{\pi \tau f} \right)^2 \sin^2 \pi T f df \quad (35)$$

In this approach, for a fixed value of τ , $I_y^2(T, \tau)$ is plotted as a function of T . The bias function B_2 studied by Barnes²⁸ is useful since it gives the link between $\langle \sigma_y^2(2, T, \tau) \rangle$ and $\langle \sigma_y^2(2, \tau, \tau) \rangle$.

For $T > \tau$: $\langle \sigma_y^2(2, T, \tau) \rangle = \frac{2}{T} \sigma_y^2(\tau)$ for white and flicker phase noises;

$\langle \sigma_y^2(2, T, \tau) \rangle = \sigma_y^2(\tau)$ for white frequency noise;

$\langle \sigma_y^2(2, T, \tau) \rangle > \sigma_y^2(\tau)$ for flicker frequency noise.

Structure function approach^{24, 25}.

Lindsey and Chie have re-interpreted oscillator instability measures in terms of the Kolmogorov²⁹ structure functions. When applied to phase and frequency fluctuations in oscillators, this approach plays a unifying role and offers a unified mathematical characterization of phase instability and frequency instability; besides, it shows the degree of sameness of these two concepts and their interconnections with other stability measures. In particular, it has been shown how $I_y^2(\tau)$ and $\sigma_y^2(\tau)$ are related to structure functions^{24, 25}. It will be shown here that other parameters such as $\sigma_{H_{pc}}^2(\tau)$, $\Sigma_y^2(\tau)$, $E[\delta y_m^2]$ and $I_y^2(T, \tau)$ may also be re-interpreted in terms of structure functions.

Other advantages of structure functions have been outlined in^{24, 25}, especially its ability to deal with problems associated with the long term frequency drifts and the flicker-type noises.

Definition of structure functions.

The N th increment, $N > 1$, of a random process $x(t)$ is defined by:

$$\Delta^N x(t; \tau) = \sum_{k=0}^{N-1} (-1)^k \binom{N}{k} x(t + (N-k)\tau) \quad (36)$$

$x(t)$ is called a process with stationary N th increment³⁰ if the following mathematical expectations:

$$E[\Delta^N x(t; \tau)] = c(\tau) \quad (37)$$

$$E[\Delta^N x(t; \tau) \Delta^N x(t+T; \tau)] = D_x^{(N)}(T; \tau) \quad (38)$$

exist for all real T and τ and do not depend on t . It means that $\Delta^N x(t; \tau)$ has a time-independent mean and its autocorrelation depends only on the time difference (wide sense stationarity).

The function $D_x^{(N)}(T; \tau)$ is the structure function of the N th increment. In the following, we will consider mainly the case where $T = 0$; $D_x^{(N)}(\tau)$ will then be referred to as the N th structure function of the process $x(t)$:

$$D_x^{(N)}(\tau) = E[(\Delta^N x(t; \tau))^2] \quad (39)$$

For oscillator stability characterization, it is interesting to apply this concept to the phase process $\varphi(t)$ first***. For a better understanding of $D_\varphi^{(N)}(\tau)$, it is interesting to re-write it in terms of the \bar{y}_i 's.

* $E[\]$ denotes the mathematical expectation operator.

** Ergodicity is needed for identity of ensemble averages with infinite time averages.

*** The N th increment of the phase process has been previously used by Barnes³³ as a powerful means of classifying the statistical fluctuations of the phase and frequency in oscillators.

Interpretation of $D_{\varphi}^{(N)}(\tau)$.

By definition, $D_{\varphi}^{(N)}(\tau)$ involves the instantaneous values of $\varphi(t)$ at $t, t+\tau, t+2\tau, \dots, t+N\tau$. The i th value $\varphi(t+(i-1)\tau)$ with $i = 1$ to $N+1$ is weighted by $(-1)^i \binom{N}{i-1}$ when N is odd and by $(-1)^{i+1} \binom{N}{i-1}$ when N is even. The two cases arise because the last term $\varphi(t+N\tau)$ has always a positive sign by definition of $\Delta^N \varphi(t; \tau)$. It is easy to show that the N th increment of $\varphi(t)$ may be rewritten in terms of the \bar{y}_i ($i = 1$ to N):

$$\Delta^N \varphi(t; \tau) = \omega_0 \tau \sum_{i=1}^N (-1)^{i+1} \binom{N-1}{i-1} \bar{y}_i \quad N \text{ odd} \quad (40)$$

$$\Delta^N \varphi(t; \tau) = \omega_0 \tau \sum_{i=1}^N (-1)^i \binom{N-1}{i-1} \bar{y}_i \quad N \text{ even} \quad (41)$$

Let us now consider the successive values of N .

Case $N = 1$.

In this case: $\Delta^1 \varphi(t; \tau) = \omega_0 \tau \bar{y}_1$
and $D_{\varphi}^{(1)}(\tau) = \omega_0^2 \tau^2 E[\bar{y}_1^2] = \omega_0^2 \tau^2 I^2(\tau)$

The true variance is hence directly related to the first structure function of the phase process^{24, 25}:

$$I(\tau) = \frac{1}{\omega_0 \tau} \sqrt{D_{\varphi}^{(1)}(\tau)} \quad (42)$$

Case $N = 2$.

$$\Delta^2 \varphi(t; \tau) = \omega_0 \tau (\bar{y}_2 - \bar{y}_1)$$

$$D_{\varphi}^{(2)}(\tau) = \omega_0^2 \tau^2 E[(\bar{y}_2 - \bar{y}_1)^2] = 2\omega_0^2 \tau^2 \sigma_y^2(\tau)$$

The two-sample zero-dead-time sample variance is simply related to the second structure function of the phase process^{24, 25}

$$\sigma_y^2(\tau) = \frac{1}{\sqrt{2} \omega_0 \tau} \sqrt{D_{\varphi}^{(2)}(\tau)} \quad (43)$$

Case $N = 3$.

$$\Delta^3 \varphi(t; \tau) = -\omega_0 \tau (2\bar{y}_2 - \bar{y}_1 - \bar{y}_3)$$

$$D_{\varphi}^{(3)}(\tau) = \omega_0^2 \tau^2 E[(2\bar{y}_2 - \bar{y}_1 - \bar{y}_3)^2]$$

A comparison with equation (29) shows that the third structure function of the phase is simply related to the three-sample zero-dead-time modified sample variance introduced in this paper following the ideas of Picinbono et al:

$$\Sigma_y(\tau) = \frac{1}{3\omega_0 \tau} \sqrt{D_{\varphi}^{(3)}(\tau)} \quad (44)$$

Case where N is even.

The N th structure function of $\varphi(t)$ reads as:

$$D_{\varphi}^{(N)}(\tau) = \omega_0^2 \tau^2 E\left[\left(\sum_{i=1}^N (-1)^i \binom{N-1}{i-1} \bar{y}_i\right)^2\right] \quad (45)$$

A look at equation (18) shows that the above mathematical expectation is nothing but the Hadamard variance with adjacent samples weighted by binomial coefficients¹⁵:

$$\langle \sigma_{H_{\varphi}}^2(N, \tau, \tau) \rangle^{1/2} = \frac{1}{\omega_0 \tau} \sqrt{D_{\varphi}^{(N)}(\tau)} \quad (46)$$

From equation (19) with $T = \tau$ and equation (46), it is easy to deduce the very important relation between $D_{\varphi}^{(N)}(\tau)$ and $S_{\varphi}(f)$:

$$D_{\varphi}^{(N)}(\tau) = 2^{2N} \int_0^{\infty} S_{\varphi}(f) \sin^{2N} \pi \tau f df \quad (47)$$

which shows clearly the interest of using higher order from the point of view of convergence for spectra with power laws.

It is important to emphasize on the fact that $\varphi(t)$ possesses a power spectral density in the usual sense only if it is stationary; if it is not, it may be convenient to define formally $S_{\varphi}(f)$ by the relationship:

$$S_{\varphi}(f) = \frac{S_{\dot{\varphi}}(f)}{4\pi^2 f^2} = \frac{S_{\omega}(f)}{f^2} \quad (48)$$

Although the Hadamard variance was explicitly defined with N even, equation (47) is valid as well for N odd and may be rewritten as:

$$D_{\varphi}^{(N)}(\tau) = 2^{2N-2} (\omega_0 \tau)^2 \int_0^{\infty} S_{\varphi}(f) \frac{\sin^{2N} \pi \tau f}{(\pi \tau f)^2} df$$

Phase instability vs frequency instability.

Although these concepts are often confused, two different definitions may be given²⁵, related to the first structure functions of phase and frequency relatively.

Phase instability.

$I_{\varphi}(\tau)$ of an oscillator is defined as the ratio of the rms value of the amount of phase noise accumulated in τ seconds to the phase accumulated if the oscillators were perfect ($\omega_0 \tau$):

$$I_{\varphi}(\tau) = \frac{\sqrt{E[(\varphi(t+\tau) - \varphi(t))^2]}}{\omega_0 \tau} = \frac{1}{\omega_0 \tau} \sqrt{D_{\varphi}^{(1)}(\tau)} \quad (49)$$

This is nothing but the classical true variance $I(\tau)$ usually presented as a measure of frequency instability (see eq. (8) and (42)).

Fractional frequency instability $I_y(T)$ of an oscillator over a time interval T is defined as the rms value of the amount of instantaneous fractional frequency deviations taking place in T seconds:

$$I_y(T) = \sqrt{E[(y(t+T) - y(t))^2]} = \sqrt{D_y^{(1)}(T)} \quad (50)$$

This parameter is nothing but the parameter $\sqrt{E[\delta y_{\text{rms}}^2]}$ given by equation (31) with $T = mT_0$. Frequency instability over T is hence defined in terms of the first structure function of frequency fluctuations, assuming that each frequency measurement is made instantaneously ($\tau = 0$) which is a serious limitation to the practical usefulness of $I_y(T)$.

For stationary frequency fluctuations, the following limits hold:

$$\lim_{T \rightarrow 0} I_y(T) = 0 \quad \lim_{T \rightarrow \infty} I_y(T) = 2E[y^2(t)] \quad (51)$$

The general behavior is hence very different from the one of $I^2(\tau)$ since:

$$\lim_{\tau \rightarrow 0} I^2(\tau) = E[y^2(t)] \quad \lim_{\tau \rightarrow \infty} I^2(\tau) = 0 \quad (52)$$

Frequency instability with $\tau \neq 0$.

A more meaningful quantity may be obtained by replacing the instantaneous values in (50) by averages over a time interval τ measured at the instants t and $t+T$. One gets then the parameter $I_y^2(T; \tau)$ described by equation (35) which is twice the sample variance with $N = 2$ and $T > \tau$. This parameter may be related to structure functions by looking at the general expression of $D_x^{(N)}(T; \tau)$ for $x = y$, $N = 1$ and $T \neq 0$ (equation (38)) :

$$D_y^{(1)}(T; \tau) = E[(y(t+\tau) - y(t)) \cdot (y(t+T+\tau) - y(t+T))] \quad (53)$$

$$= \omega_p^2 \tau^2 E[\bar{y}_1 \cdot \bar{y}_2]$$

Since :

$$I_y^2(T; \tau) = E[(\bar{y}_1 - \bar{y}_2)^2] = 2 \{ E[\bar{y}_1^2] - E[\bar{y}_1 \bar{y}_2] \} \quad (54)$$

$$I_y^2(T; \tau) = \frac{2}{\omega_p^2 \tau^2} \{ D_y^{(1)}(\tau) - D_y^{(1)}(T; \tau) \} \quad (55)$$

Conclusion

The various parameters that have been proposed by several authors as frequency/phase stability measures have been reviewed. The key role played by the spectral density of frequency fluctuations appeared from the numerous equations involving a transfer function. The knowledge of its transfer function appears thus as one of the best means to study the properties of a time domain parameter.

While reviewing the various approaches, some new parameters have been considered and some relations between time domain parameters and structure functions have been established.

In the method of the finite-time frequency control, a specific parameter $\Sigma_y(\tau)$ has been defined and its relationship to the 3rd structure function of phase noise has been demonstrated.

The link between $E[\delta y_{\text{rms}}^2]$ and the first structure function of frequency noise has been shown, and the more realistic case of non-instantaneous measurements has been treated with its relation to structure functions and to sample variances.

Also, the Hadamard variance weighted by binomial coefficients has been linked to the Nth order structure function of phase noise. The high-pass and bandpass variances which are defined by their transfer functions and measured by analog filtering techniques are not directly linked to any structure function.

The structure function approach appears as a very general mean of introducing the main time domain parameters measured with counters. On the other hand, recognizing the importance of the transfer function, the transfer function approach appears as another interesting mean for defining and measuring time domain parameters without counters.

It is hoped that this paper will be useful for people faced with the problem of determining the characteristics of oscillators. Also, some of the concepts presented here or other improvements³¹ may become of primary importance in the future with the advent of new sources and/or applications.

References

- 1 "IEEE-NASA Symposium on short-term frequency stability". Goddard Space Flight Center, Greenbelt Maryland, Nov. 1964.
- 2 "Proc. of the IEEE, Special issue on frequency stability". Vol 54, N° 2, Feb. 1966, pp. 101-338.
- 3 Baghdady E. J. et al : "Short Term Frequency Stability : Characterization, Theory and Measurement". Ref. 1, pp. 65-87.
Cutler L. S. : "Some aspects of the theory and measurement of frequency fluctuations in frequency standards". Ref. 1, pp. 89-100.
- 4 Cutler L. S., Searle C. L. : "Some aspects of the theory and measurement of frequency fluctuations in frequency standards". Ref. 2, pp. 136-154.
- 5 Allan D. W. : "Statistics of atomic frequency standards", Ref. 2, pp. 221-230.
- 6 Barnes J. A. et al : "Characterization of Frequency Stability". NBS Technical Note 394, October 1970.
- 7 Barnes J. A. et al : "Characterization of Frequency Stability". IEEE Tr. Instr. Meas., Vol IM-20, N° 2, May 1971, pp. 105-120.
- 8 See for example the "Proceedings of the Annual Symposium on Frequency Control" or the "Proceedings of the Annual Precise Time and Time Interval (PTTI) Planning Meeting".
- 9 Slepian D. : "On bandwidth". Proc. of the IEEE, Vol. 64, N° 3, March 1976, pp. 292-300.
- 10 Boileau E., Picinbono B. : "Statistical Study of phase fluctuations and oscillator stability". IEEE Tr. Instr. Meas., Vol. 25, N° 1, March 1976, pp. 66-75.
- 11 Papoulis A. : "Probability, Random Variables and Stochastic Processes". Mc Graw Hill Book Company, 1965.
- 12 Blaqui re A. : "Effet du bruit de fond sur la fr quence des auto-oscillateurs   lampes. Pr cision ultime des horloges radio lectriques". Ann. de Radio lectricit , tome VIII, N° 31, jan. 1953, pp. 36-80.
- 13 Edson W. A. : "Noise in Oscillators". Proc. of the IRE, Vol. 48, N° 8, August 1960, pp. 1454-1466.
- 14 Lamb W. E. : "Theory of an optical maser". Phys. Rev., Vol. 134-A, June 1964, pp. 1429-1450.
- 15 Baugh R. A. : "Frequency Modulation analysis with the Hadamard variance". Proc. of the 25th annual symposium on Frequency Control, April 1971, pp. 222-225.

* (equations (9), (10), (12), (14), (16), (19), (21), (23), (26), (28), (30), (32), (33), (35), (47)).

- 16 Lesage P., Audoin C. : "Characterization of frequency stability : uncertainty due to the finite number of measurements". IEEE Tr. Meas., Vol. IM-22, N° 2, June 1973, pp. 157-161.
- 17 Rutman J. : "Comment on : Characterization of frequency stability". IEEE Tr. Instr. Meas., Vol. IM 21, N° 2, February 1972, p. 85.
- 18 Rutman J. : "Instabilité de fréquence des Oscillateurs". L'Onde Electrique Vol 52, N° 11, December 1972, pp. 480-487.
- 19 Sauvage G., Rutman J. : "Analyse spectrale du bruit de fréquence des oscillateurs par la variance de Hadamard". Ann. des Télécommunications, Tome 28, N° 7-8, juillet-août 1973, pp. 304-314.
- 20 Lesage P., Audoin C. : "Instabilité de fréquence des oscillateurs : mesure dans le domaine fréquentiel par une méthode temporelle". L'onde électrique, Vol 55, N° 2, Février 1975, pp. 82-89.
- 21 Gros Lambert J. : "A new spectrum analyzer using both analog and digital filtering via Hadamard variance". IEEE Tr. Instr. Meas., Vol IM-25, N° 4, December 1976, pp. 480-483.
- 22 Rutman J. : "Characterization of frequency stability : a transfer function approach and its application to measurements via filtering of phase noise". IEEE Tr. Instr. Meas., Vol IM-23, N° 1, March 1974, pp. 40-48.
- 23 De Prins J., Cornelissen G. : "Power Spectrum, Frequency Stability and Flicker Noise". Proc. of the Frequency Standards and Metrology Seminar, Univ. Laval, Quebec, Canada, 30-31 Aug.-1 Sept 1971, pp. 368-387.
- 24 Lindsey W.C., Chie C.M. : "Interpretation and application of oscillator instability measures using structure functions". Proc. of the 8th Annual PTTI Applications and Planning Meeting, Nov. 30-December 2, 1976.
- 25 Lindsey W.C., Chie C.M. : "Theory of oscillator instability based upon structure functions". Proc. of the IEEE, Vol 64, N° 12, Dec. 1976, pp. 1652-1666.
- 26 Rutman J., Sauvage G. : "Measurement of frequency stability in time and frequency domains via filtering of phase noise". IEEE Tr. Instr. Meas., Vol IM-23, N° 4, Dec. 1974, pp. 515-518.
- 27 Cornelissen G. : "Analyse de signaux et application aux problèmes de définition de la stabilité de fréquence". Université libre de Bruxelles, Thèse de Doctorat ès Sciences Physiques, Année Académique 1975-1976.
- 28 Barnes J.A. : "Tables of bias functions, B_1 and B_2 , for variances based on finite samples of processes with power law spectral densities". NBS Technical Note 375, January 1969.
- 29 Kolmogorov A.N. : "Dissipation of energy in locally isotropic turbulence". Dokl. Nauk. SSR, Vol 32, 1941, pp. 16-18.
- 30 Yaglom A.M. : "Correlation theory of processes with random stationary Nth increments". Ann. Math. Soc. Trans., Vol 8, 1958, pp. 87-141.
- 31 Boileau E. : "Improvements to the procedures used to study the fluctuations of oscillators". Submitted for publication in the IEEE Tr. Instr. Meas.
- 32 Barnes J.A. : "Models for the interpretation of frequency stability measurements". NBS Technical Note 683, August 1976.
- 33 Barnes J.A. : "Atomic timekeeping and the statistics of precision signal generators". Ref. 2, pp. 207-220.
- 34 Wiley R.G. : "A direct time-domain measure of frequency stability : the modified Allan variance". IEEE Tr. Instr. Meas., Vol IM-26, N° 1, March 1977, pp. 38-41.
- 35 Fischer M.C. : "Frequency stability measurement procedures". Proc. of the 8th Annual PTTI Applications and Planning Meeting, Nov. 30 - Dec. 2, 1976.
- 36 Winkler G.M.R. : "A brief review of frequency stability measures". Proc. of the 8th Annual PTTI Applications and Planning Meeting, Nov 30 - Dec. 2, 1976.

Note added in proof : References 35 and 36 have been received lately. Practical questions related to measurement techniques are discussed in³⁵. In³⁶, emphasis is put on the dichotomy between trends, systematic or deterministic variations of frequency and random variations of frequency. The necessity to remove any obvious systematics before the random part of the clock performance is analyzed is also emphasized. A so-called "curvature variance" used by Kramer at the PTB is also discussed : it is equal to 1.5 times the $\sum_y^2(\tau)$ defined in this paper (equation (29)).

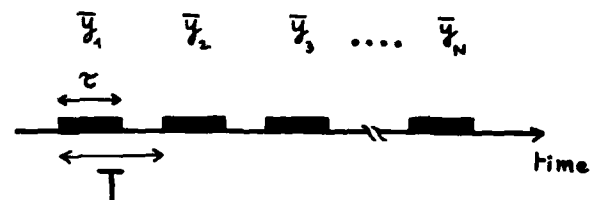


FIG. 1 Measurement cycle in the time domain

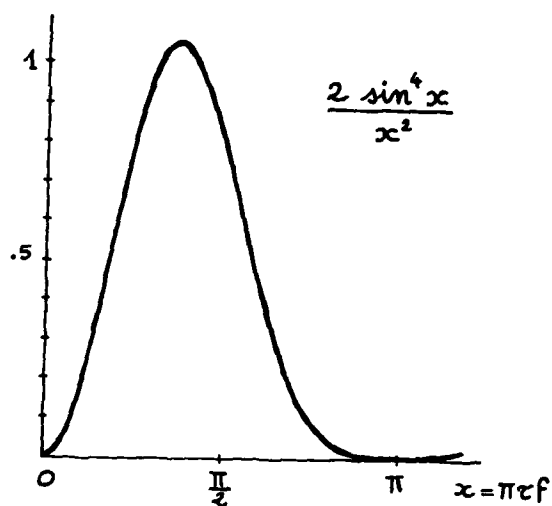
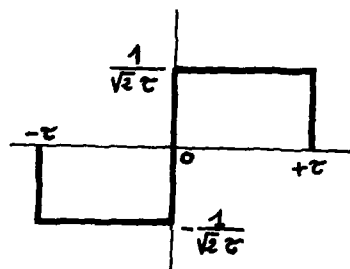


FIG. 2 Allan Variance: a) Measurement Sequence
b) Transfer Function

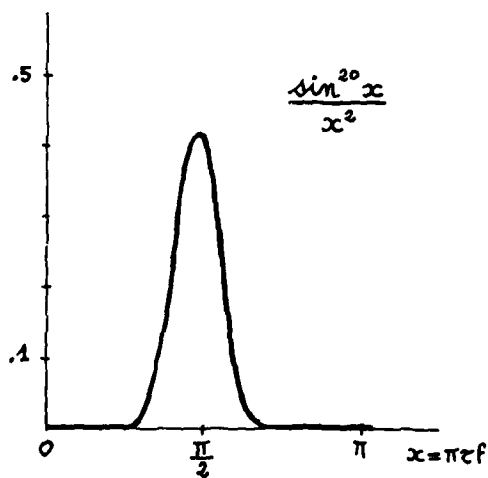


FIG. 4 Transfer Function of the Hadamard Variance
Weighted by Binomial Coefficients ($N=10$),
Divided by 2^{18} .

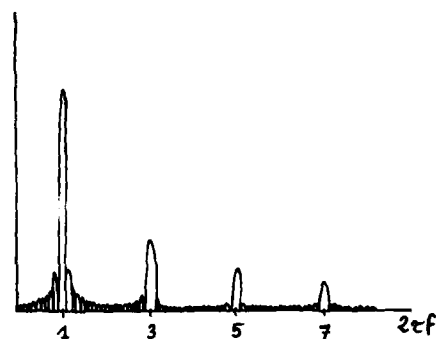
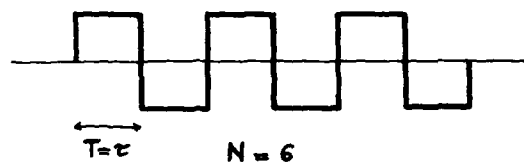


FIG. 3 Hadamard Variance: a) Measurement Sequence
($N=6, T=\tau$).
b) Shape of the Transfer
Function

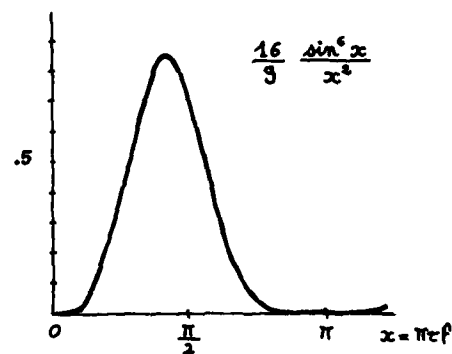
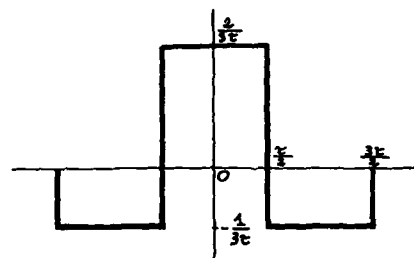


FIG. 5 Modified Three-sample Variance:
a) Measurement Sequence
b) Transfer Function

SPECIFICATION AND MEASUREMENT OF OSCILLATOR PHASE NOISE INSTABILITY

W. C. Lindsey
University of Southern California
Los Angeles, CA 90007

C. M. Chie
LinCom Corporation
Pasadena, CA 91105

ABSTRACT

A current problem of interest in the telecommunication, Doppler measuring, radar and ranging system design, and procurement is that of specifying system phase noise characteristics. In order to write a meaningful phase noise specification the engineer needs a theory upon which system performance measures can be evaluated and tradeoffs made as functions of the phase noise model parameters. In addition, the phase noise model parameters must be verified by direct laboratory measurement.

One of the main purposes of this paper is to demonstrate that the performance measures of telecommunication and Doppler measuring systems are found by appropriately weighting the PSD $S_y(f)$. We note that the frequency range of interest in $S_y(f)$ depends upon the system application and performance measure selected; this fact will be demonstrated.

The paper will begin by reviewing the role which structure functions have in the theory of oscillator phase noise. We then turn to study the effect of phase noise on various applications where oscillators are involved using these functions. We shall analyze a particular measurement scheme where the test oscillator is phase-locked to a voltage controlled crystal oscillator, the so-called two oscillator test set-up. The effect of phase noise on the performance of ranging, Doppler and coherent communication systems as well as various modulation and code generation schemes will also be considered in terms of structure functions associated with the oscillator instability.

With the above discussion in mind, we shall address the problem of specification of oscillator instability for different applications. Since the τ -domain and f -domain can be interrelated, the τ -domain and f -domain specification and measurements are actually equivalent. We shall demonstrate this point by giving examples where the Allan variance is known and we calculate the rms fractional frequency deviation and the spectral density from such data.

1. INTRODUCTION

A current problem of great interest in telecommunication, Doppler measuring, radar and ranging system design, and procurement is that of specifying system phase noise requirements¹. In order to

¹By phase noise we mean any distortion which destroys the line-width of the created microwave signal by the process of phase modulation. There are various components, e.g., flicker noise from the source, media effects, AM to PM conversion from TWT amplifiers, etc.

write a meaningful phase noise specification the engineer needs a theory upon which system performance measures can be evaluated and tradeoffs made as functions of the phase noise model parameters. In addition, phase noise model parameters must be verified by direct measurement of the underlying disturbances. Unfortunately, phase noise, unlike thermal noise which is well approximated by a delta correlated process, does not possess a constant power spectral density for all frequencies; not only does this fact make it difficult to measure but makes it more difficult to develop a working theory from which performance can be predicted [1], [2].

It is interesting to note that in order to determine the effects which thermal noise has on a communication link power budget (useful in predicting system margins), one only needs the system noise temperature from which the noise power spectral density N_0 watts/Hz can be determined. However, when phase noise is assumed to be present in the system model, the state of affairs is quite different; in this case the engineer also needs the power spectral density of the frequency instabilities say, $S_y(f)$ [2].

The paper will begin by reviewing the role which structure functions have in the theory of oscillator phase noise [1], [2]. Interconnections of phase structure functions and the present 'standards' will be pointed out. A commonly encountered example of an oscillator with frequency drift and power-law type PSD will be given to demonstrate the interconnections.

Once this notion is clarified, we turn to study the effect of phase noise on various applications where oscillators are involved using these functions and the filter approach. One immediate application is in laboratory phase noise measurements. We shall consider a particular measurement scheme where the test oscillator is phase-locked to a voltage controlled crystal oscillator, the so-called two oscillator test set-up. What we observe in such a set-up is a filtered version of the phase noise process; the filter being the phase locked loop. In addition, if we work with increments of the phase noise process, we are, in effect, passing the phase noise through a digital filter. We shall give a detailed discussion in this paper of the effects of the closed loop transfer function on the measurement and how the measurement can be reflected into the true PSD of the phase noise process. The effect of phase noise on the performance of ranging, Doppler and coherent communication systems as well as various modulation and code generation schemes will also be considered in terms of structure functions associated with the oscillator instability.

With the above discussion in mind, we shall address the problem of specification and measurement of oscillator instability. For several applications the PSD of the frequency noise process is the

most convenient form for use in specification work. There are other reasons for such a choice. In particular, we shall show how frequency multiplication affects the RF power spectral linewidth. This information cannot be directly obtained from τ -domain specification such as the Allan variance. The less convenient τ -domain measures which are, however, useful in a particular application, should be specified in terms of the two sets of stability functions of the filtered frequency and phase noise process. We have seen this in clock application, viz., the rms fractional frequency deviation and the Allan variance.

Finally, we shall bear in mind that the τ -domain and f-domain can be interrelated. Hence the τ -domain and f-domain specification and measurements are equivalent. The only point to make is that the PSD is what is needed in writing specifications for a host of applications other than precise time-keeping. We shall demonstrate this point by giving a numerical example where the Allan variance is known and we calculate the rms fractional frequency deviation. We also show how the spectral density can be computed from such data.

II. STRUCTURE FUNCTIONS IN THE THEORY OF OSCILLATOR PHASE NOISE

A. Structure Function

We begin by reviewing the role which structure functions have in the theory of oscillator phase noise and by setting the "standard" notation [3] adopted in this paper. The instantaneous output voltage of a high-quality signal generator may be written as [3]

$$V(t) = [V_0 + \epsilon(t)] \sin[2\pi\nu_0 t + \psi(t)] \quad (1)$$

V_0 and ν_0 are the nominal amplitude and frequency, respectively, ($\omega_0 = 2\pi\nu_0$). $\epsilon(t)$ and $\psi(t)$ are random processes representing amplitude noise and phase noise, respectively. The random process $\dot{\psi}(t) = d\psi(t)/dt$ denotes angular frequency fluctuations and is often assumed to have a zero mean value. Provided $\epsilon(t)$ and $\psi(t)$ are sufficiently small for all time t , the fractional instantaneous frequency deviation from nominal is defined as

$$y(t) = \frac{1}{2\pi\nu_0} \frac{d\psi(t)}{dt} = \frac{\dot{\psi}(t)}{\omega_0} \triangleq \dot{x}(t) \quad (2)$$

The recommended definition for the frequency stability measure in the Fourier frequency-domain is the one-sided spectral density on a per Hertz basis $S_y(f)$ of the pure real random process $y(t)$. Denoting $S_{\dot{\psi}}^y(f)$ the spectral density of angular frequency fluctuations, then

$$S_y(f) = \frac{1}{2} S_{\dot{\psi}}^y(f) \quad (3)$$

For most oscillators, the assumption that $y(t)$ has a zero mean value is not always valid. As an example, state of the art crystal oscillators exhibit a typical frequency drift component $E[y(t)] = d_1 t$ where d_1 is of the order of $5 \times 10^{-10}/\text{day}$ [4]. From this point on we assume that the frequency process $y(t)$ has a time-dependent mean drift (possibly zero),

viz.,

$$y(t) = E[y(t)] + \tilde{y}(t) \quad (4)$$

where $\tilde{y}(t)$ is a zero mean random process and

$$E[y(t)] = \sum_{k=1}^{N-1} \frac{d_k}{k!} t^k \quad (5)$$

Let us denote the PSD of the process $\tilde{y}(t)$ by $S_{\tilde{y}}(f)$ (instead of $S_y(f)$) to comply with the "standard" notation [3]. Notice that if $d_1 = \dots = d_{N-1} = 0$ in (5) then $\tilde{y}(t) = y(t)$ and our generalization reduces to the conventional definition.

It is well known that taking higher-order increments reduces the degree of a polynomial. In particular, the $(N-1)^{\text{st}}$ increment of $y(t)$ reduces to [1] ($\underline{\tau}^m$ is the m -tuple (τ_1, \dots, τ_m))

$$\Delta^{N-1} y(t; \underline{\tau}^{N-1}) = d_{N-1} \prod_{k=1}^{N-1} \tau_k + \Delta^{N-1} \tilde{y}(t; \underline{\tau}^{N-1}) \quad (6)$$

and the N^{th} increment of $y(t)$ reduces to

$$\Delta^N y(t; \underline{\tau}^N) = \Delta^N \tilde{y}(t; \underline{\tau}^N) \quad (7)$$

which is independent of the systematic drift. Let us define the m^{th} structure function of a process $\xi(t)$ by [1] (assuming stationarity)

$$D_{\xi}^{(m)}(\underline{\tau}^m) \triangleq E\{\left[\Delta^m \xi(t; \underline{\tau}^m)\right]^2\} \quad (8)$$

For $m = N$, we observe that $D_y^{(m)}(\underline{\tau}^m) = D_{\tilde{y}}^{(m)}(\underline{\tau}^m)$ which is independent of the systematic drift. In [1], it was shown that $D_y^{(M)}(\underline{\tau}^M)$ and the PSD $S_y(f)$ are related via

$$D_y^{(M)}(\underline{\tau}) = 4^M \int_0^\infty \prod_{k=1}^M (\sin^2 \pi \tau_k f) S_y(f) df, \quad M \geq N \quad (9)$$

Similar results are developed [1] for the fractional phase process $x(t)$. In particular, it was shown that [1]

$$\Delta^N x(t; \underline{\tau}^N) = d_{N-1} \prod_{k=1}^N \tau_k + \Delta^{N-1} \left[\int_t^{t+\tau} \tilde{y}(\xi) d\xi \right] \quad (10)$$

$$\Delta^{M+1} x(t; \underline{\tau}^{M+1}) = \Delta^M \left[\int_t^{t+\tau} \tilde{y}(\xi) d\xi \right]; \quad M \geq N \quad (11)$$

and

$$D_x^{(M+1)}(\underline{\tau}^{M+1}) = 4^{M+1} \int_0^\infty \prod_{k=1}^{M+1} \frac{\sin^2 \pi \tau_k f}{(2\pi f)^2} S_y(f) df; \quad M \geq N \quad (12)$$

Notice also that $D_x^{(M+1)}(\underline{\tau}^{M+1})$ is independent of the system drift. For $N=1, 2$, the structure function of the fractional phase has long been recognized, namely,

$$D_x^{(1)}(\tau) = \tau^2 \left[\frac{\Delta f(\tau)}{f_0} \right]^2 \quad (13)$$

and for (the single argument τ denotes $\underline{\tau}^k = (\tau, \dots, \tau)$)

$$D_x^{(2)}(\tau) = 2\tau^2 \sigma_y^2(\tau) \quad (14)$$

where $\Delta f(\tau)/f_0$ and $\sigma_y^2(\tau)$ denotes the rms fractional

deviation and the two-sample Allan variance respectively [3].

In practice, the PSD $S_y(f)$ is usually modeled by a linear combination of power-law PSD's having the form

$$S_y(f) = \begin{cases} h_v f^v & 0 \leq f \leq f_h \\ 0 & f > f_h \end{cases} \quad (15)$$

Using this model of $S_y(f)$ in (9) and (12), we arrive at two important relationships, namely,

$$D_x^{(M+1)}(\tau) = \frac{4^{M_h} h_v}{\pi^{v+1}} F(v-2, M+1, \pi \tau f_h) \tau^{1-v}; M \geq N \quad (16)$$

and

$$D_y^{(M)}(\tau) = \frac{4^{M_h} h_v}{\pi^{v+1}} F(v, M, \pi \tau f_h) \tau^{-v-1}; M \geq N \quad (17)$$

where (μ not necessarily an integer)

$$F(\mu, m, B) \triangleq \int_0^B x^\mu \sin^2 m x dx; \mu > -(2m+1) \quad (18)$$

The function $F(\mu, m, B)$ is investigated and tabulated in [5]. For $\mu > -1$, $F(\mu, m, B)$ is dependent on B . However, for $-(2m+1) < \mu < -1$, $F(\mu, m, B)$ is essentially independent of B for $B \gg 1$. The function $F(\mu, n, \infty)$ for $-(2m+1) < \mu < -1$ is plotted in Fig. 1. Using this curve and the relations (16) and (17), one can determine $S_y(f)$ from $D_x^{(N+1)}(\tau)$ or $D_y^{(N)}(\tau)$ which are readily measured experimentally. Hence, the use of higher-order structure functions alleviates the usual convergence problem associated with the power law spectrum.

B. An Example Illustrating the Application of Structure Functions

Consider a crystal oscillator with an aging rate of $10^{-10}/\text{day}$, i.e., $N=2$, $d_1 = 1.2 \times 10^{-15}/\text{sec}$ in (14). Notice that from (10) and (12), the systematic drift gives rise to a τ^2 dependence of $\sigma_y^2(\tau)$, namely,

$$\sigma_y^2(\tau) = \frac{1}{2} d_1^2 \tau^2 + \frac{4}{\pi} \int_0^\infty \frac{\sin^4(\pi \tau f)}{f^2} S_y(f) df \quad (19)$$

If $S_y(f)$ is modeled by a flicker frequency PSD $S_y(f) = h_{-1} f^{-1}$ with $h_{-1} = 7.2 \times 10^{-25}$, $\sigma_y(\tau)$ can be depicted as in Fig. 2 [6, Fig. 16]. Notice, however, that the third difference of $x(t)$ is independent of the drift and $D_x^{(3)}(\tau) = 4.86 \times 10^{-24} \tau^2$. The normalized quantity $\sqrt{D_x^{(3)}(\tau)/\tau^2} = 2.2 \times 10^{-12}$ is a constant independent of τ . For comparison purposes, it is plotted alongside with $\sigma_y(\tau)$ in Fig. 2 (the dashed line). In addition, if the PSD has an additional independent component say, $S_y(f) = h_{-4} f^{-4}$, from (16) we conclude that [5]

$$D_x^{(3)}(\tau) = \frac{22}{5} \pi^4 h_{-4} \tau^5 \quad (20)$$

for this particular component. Thus higher-order structure function provide a general characterization of oscillator stability.

III. PHASE NOISE MEASUREMENTS USING A GENERAL LINEAR FILTER APPROACH

A. Linear Model

All physical processes can only be observed indirectly, i.e., through a filter. As an example, the power spectrum of a process indicated by a spectrum analyzer is a display of the energy distribution of the outputs of a bandpassed version of the process at the appropriate bands. The phase noise process is no exception. The general situation is depicted in Fig. 3. The phase noise² characteristics are determined by observing it through a filter $G_0(j2\pi f)$ which has inherent lower and upper cutoff frequency f_l and f_h . In a similar fashion, when phase noise degrades system performance, the phase noise process is also being filtered by a "system" filter $G_1(j2\pi f)$ with lower and upper cutoff frequency F_l and F_h . This arrangement is indicated in Fig. 4. From these two models, we see that what we are really dealing with, i.e., what can be measured and what affects system performance, is not the phase noise process directly, but a filtered version of the phase noise process. Since we are not interested in phase noise per se, we need a characterization of the phase noise which is compatible with this linear filter approach: the PSD. Before we proceed to give important examples (which includes the measurement aspects and system applications of interest) that fits in naturally with this framework, we want to make the following comments. First of all, since the filter possesses high and low cutoff frequencies, f_l^1 and f_h^1 of the phase noise is not going to affect appreciably the PSD of the filtered phase noise (hence all second moment statistics) as long as $f_l \ll f_l^1$ and $f_h \ll f_h^1$. If it turns out to be a convenient model and simplify computation, we can idealize $f_l = 0$ and $f_h = \infty$. It is only in this context that we allow an infinite phase noise variance, as the case of white or flicker noise. Otherwise, we must not allow infinite variance and appropriately specify f_l^1 and f_h^1 . Secondly, we note that taking the N^{th} increment of the phase noise is equivalent to passing it through a filter with³

$$G_2(j2\pi f) = \prod_{k=1}^N (e^{j2\pi f \tau_{k-1}}). \quad \text{This effect lumps with the measurement filter or the system filter.}$$

B. Two-Oscillator Phase Noise Measurement

The functional diagram and the linear model for the special case of a two-oscillator phase noise test set-up is shown in Fig. 5 [7]. The loop phase error is given by (p is the Heaviside operator)

²For simplicity of notation, we take phase noise to denote the frequency noise $y(t)$, if we are interested in $x(t)$, a factor of $1/j2\pi f$ can be included in $G_0(j2\pi f)$.

³The advance operator is due to the use of the forward difference definition, for a real system an appropriate time scale adjustment is implied.

$$\varphi = \psi_1 - \left[\psi_2 + \frac{AKF(p)}{p} \varphi \right] \quad (21)$$

where A^2 is the rms power of the oscillation under test, K is the loop gain, $F(p)$ is the loop filter, and ψ_1, ψ_2 are the phase noise of the test and reference oscillator, respectively. In terms of the loop transfer function $H(p) \triangleq AKF(p)/[p+AKF(p)]$, we have

$$\varphi = [1-H(p)] (\psi_1 - \psi_2) \quad (22)$$

Let us examine (22) more carefully. For a perfect $(N+1)^{st}$ order loop, the N -pole loop filter $F_N(p)$ has the form

$$F_N(p) = 1 + \sum_{k=1}^N \frac{a_k}{p^k} \quad (23)$$

from which we find $H_{N+1}(p)$ is the corresponding closed-loop transfer function for $F_N(p)$

$$1 - H_{N+1}(p) = \frac{p^{N+1}}{p^{N+1} + AKp^N + AKa_1p^{N-1} + \dots + AKa_N} \quad (24)$$

By considering the steady state behavior of (22), we see that the $(N+1)^{st}$ order loop is capable of tracking out a N^{th} order polynomial mean phase drift in ψ_1 or ψ_2 . Assuming that this is done, we can now let $E[\psi_1] = E[\psi_2] = 0$. In the case that the reference phase noise ψ_2 is negligible when compared to the phase noise contribution ψ_1 of the test oscillator,

$$\varphi = [1-H_{N+1}(p)] \psi_1 \quad (25)$$

Hence, what we measured in such a set-up is really a filtered version of ψ_1 , namely, by $[1-H_{N+1}(p)]$. It is convenient to define the power transfer function to be the ratio of the observed noise spectrum to the original noise spectrum, which is presently

$$P(f) = \frac{S_\varphi(f)}{S_{\psi_1}(f)} = |1-H_{N+1}(j2\pi f)|^2 \quad (26)$$

As an example, let us consider the first-order loop with $F_0(p) = 1$. Using $B_L \triangleq AK/4$ for the RMS loop bandwidth, the power transfer function $P_1(f)$ for a first-order loop is

$$P_1(f) = \frac{(f/4B_L)^2}{1 + (f/4B_L)^2} \quad (27)$$

which is plotted in Fig. 6(a). If the phase noise has a white frequency spectrum, i.e., $S_{\psi_1}(f) = h_0/f^2$, then the observed φ process has a PSD

$$S_\varphi(f) = \frac{h_0}{16B_L^2 [1 + (f/4B_L)^2]} \quad (28)$$

which has a finite variance $\sigma_\varphi^2 = h_0 \pi / 8B_L$. The PSD of $S_\varphi(f)$ is shown in Fig. 6(b).

For τ -domain measurements, the increment

$$\Delta^M \varphi(\tau^M) = \left[\prod_{k=1}^M (e^{j p \tau_{k-1}}) \right] [1-H_{N+1}(p)] \psi_1 \quad (29)$$

is of interest. The analogous power transfer function for the M^{th} phase error increment is therefore

$$\hat{P}_{N+1}^M(f) = 4^M |1-H_{N+1}(j2\pi f)|^2 \prod_{k=1}^M \sin^2(\pi \tau_k f) \quad (30)$$

where $H_{N+1}(j2\pi f)$ is the closed loop transfer function of an $(N+1)$ -th order loop. For the case of the first-order loop described above, (30) reduces to

$$\hat{P}_{N+1}^M(f) = 4^M \prod_{k=1}^M \sin^2 \left(4\pi B_L \tau_k \frac{f}{4B_L} \right) \frac{(f/4B_L)^2}{1 + (f/4B_L)^2} \quad (31)$$

For $M=1$ (first-order increment (31) is depicted in Fig. 6(c) for $B_L \tau = 1$. The PSD $S_{\Delta\varphi}(f)$, when $S_{\psi_1}(f) = h_0/f^2$, is given by

$$S_{\Delta\varphi}(f) = \frac{h_0}{4B_L^2} \frac{\sin^2 \left(4\pi B_L \tau \frac{f}{4B_L} \right)}{1 + (f/4B_L)^2} \quad (32)$$

and is depicted in Fig. 6(d).

Now let us quantify the ability of the two oscillator measurement technique to measure the phase noise PSD. Referring back to the linear model of Fig. 5(b), we see that the output phase θ of the VCO is given by

$$\theta = [1-H(p)] \psi_1 + H(p) \psi_2 \quad (33a)$$

Then

$$S_\theta(f) = |1-H(j2\pi f)|^2 S_{\psi_1}(f) + |H(j2\pi f)|^2 S_{\psi_2}(f) \quad (33b)$$

Hence for $f < B_L$, the loop output has a phase noise PSD of approximately $S_{\psi_1}(f)$ and for $f > B_L$, the loop output has a phase noise PSD of approximately $S_{\psi_2}(f)$. If oscillator 1 has lower phase noise PSD than oscillator 2 for $f < B_L$ but higher for $f > B_L$, e.g., oscillator 2 is a temperature-compensated crystal oscillator and oscillator 1 is an atomic standard, then the loop output oscillation has a considerably better performance than either oscillator 1 or oscillator 2.

C. Effect of Phase Noise on System Performance

The effect of phase noise on the performance of ranging, Doppler and coherent communications as well as various modulation and code generation schemes has been considered by the authors in a previous paper [2]. Using the results in [2] as illustrating examples, we attempt to merely present herein the novel features of our approach.

Consider the general situation depicted in Fig. 7. The input signal to the system of interest is corrupted by phase noise. The desired output from

the system is represented as increments of the signal and phase noise, modified by the system. The system effect is modeled by the linear filter $G(p)$. Hence the noise variance on the desired output is

$$\mathbb{D}_x^{(m)}(\tau^m) = 4 \int_0^m \sum_{k=1}^{\infty} \sin^2(\pi \tau_k f) |G(j2\pi f)|^2 \cdot \frac{S_y(f)}{(2\pi f)^2} df \quad (34a)$$

if the signal is in the phase process, and

$$\mathbb{D}_y^{(m)}(\tau^m) = 4 \int_0^m \sum_{k=1}^{\infty} \sin^2(\pi \tau_k f) |G(j2\pi f)|^2 S_y(f) df \quad (34b)$$

if the signal is in the frequency process. Notice how in (34a) and (34b), the phase noise PSD is weighted by the filter $G(j2\pi f)$ and $\prod_{k=1}^m \sin(\pi \tau_k f)$.

In almost all system applications, it is always possible to specify the performance measures in terms of (34a) or (34b). As illustrating examples, the results derived in [2] are summarized, along this guideline, in Table I.

IV. SPECIFICATION AND MEASUREMENT OF OSCILLATOR INSTABILITY

A. PSD as the Key to Specify Oscillator Instability

In the previous section, we have demonstrated through a particular measurement set-up, how oscillator instability is measured and how the measurement system affects the observed phase noise. We have also surveyed the effects of phase noise on the performance of tracking, Doppler, ranging and coherent communication systems as well as various modulation and code generation schemes. It should now become clear that phase noise effects are not isolatable from, but must be studied within the scope of the particular application in mind. In order to be useful, any specification of oscillator instability must be sufficiently versatile and general enough to handle various applications. A logical candidate is the f -domain characterization $S_y(f)$. The rationale behind this choice is clear: In all applications so far encountered, the desired signal is always corrupted by increments of the filtered phase noise. The variance of this quantity is the structure function of the filtered phase noise process which can be conveniently expressed as a weighted integral of the PSD $S_y(f)$. Such an integral can be readily evaluated either analytically or numerically, depending on the complexity of the integrand. There are other reasons for such a choice. In certain applications, the effects of phase noise can only be studied in the f -domain. In particular, we shall presently show how frequency multiplication affects the RF power spectral bandwidth in terms of the convolution products of $S_y(f)$ [8]. Because of its nature, this effect cannot be studied in the τ -domain.

Let us consider the oscillation model

$$r(t) = \sqrt{2P} \cos(2\pi f_0 t + \psi(t) + \theta_0) \quad (35)$$

where

$$\psi(t) = \int_{-\infty}^t \dot{\psi}(\tau) d\tau \quad (36)$$

Let us assume further that $\dot{\psi}(t)$ is Gaussian with a PSD $S_{\dot{\psi}}(f)$. If we make the usual assumption that θ_0 is uniformly distributed over $[0, 2\pi]$, then $r(t)$ is wide-sense stationary, zero-mean with a correlation function

$$R_r(\tau) = P \cos 2\pi f_0 \tau \exp[-D_{\dot{\psi}}^{(1)}(\tau)/2] \quad (37)$$

Now the PSD of oscillation $r(t)$ is given by

$$S_r(f) = P[S_{r_0}(f-f_0) + S_{r_0}(f+f_0)] \quad (38)$$

where (the symbol $*$ denotes convolution)

$$S_{r_0}(f) = \exp(-\sigma_{\dot{\psi}}^2) \left[\delta(f) + \frac{S_{\dot{\psi}}(f)}{(2\pi f)^2} + \frac{1}{2!} \frac{S_{\dot{\psi}}(f)}{(2\pi f)^2} * \frac{S_{\dot{\psi}}(f)}{(2\pi f)^2} + \dots \right] \quad (39)$$

and

$$\sigma_{\dot{\psi}}^2 \triangleq \int_0^{\infty} \frac{S_{\dot{\psi}}(f)}{(2\pi f)^2} df \quad (40)$$

is assumed to be bounded. Obviously, the delta function in (39) corresponds to unmodulated carrier and the remaining terms are due to oscillator instability.

Frequency multiplication is actually a phase angle multiplication by a factor of, say, M . The PSD of the resultant oscillation $\hat{r}(t) = \sqrt{2P} \cos[M(2\pi f_0 t + \psi(t) + \theta_0)]$

$$S_{\hat{r}}(f) = P[S_{\hat{r}_0}(f-Mf_0) + S_{\hat{r}_0}(f+Mf_0)] \quad (41)$$

where

$$S_{\hat{r}_0}(f) = \exp(-M^2 \sigma_{\dot{\psi}}^2) \left[\delta(f) + \frac{M^2 S_{\dot{\psi}}(f)}{(2\pi f)^2} + \frac{M^4}{2!} \frac{S_{\dot{\psi}}(f)}{(2\pi f)^2} * \frac{S_{\dot{\psi}}(f)}{(2\pi f)^2} + \dots \right] \quad (42)$$

If we take the first two terms in (42) for $S_{\hat{r}_0}(f)$ the residual energy $\hat{\rho}(\sigma_{\dot{\psi}})$ in the convolved phase noise spectrum is

$$\hat{\rho}(\sigma_{\dot{\psi}}) = 1 - \exp(-M^2 \sigma_{\dot{\psi}}^2) (1 + M^2 \sigma_{\dot{\psi}}^2) \quad (43)$$

Where M is large, the first-order approximation is no longer accurate and spectrum spreading (which tends to a Gaussian for high gains) is always present.

B. Specification of Oscillator Stability in the τ -Domain

The less convenient τ -domain measures which are useful in a particular application could be

specified in terms of the two sets of stability⁴ functions $\{D_x^{(k)}(\tau^k), k=1, 2, \dots\}$ and $\{D_y^{(k)}(\tau^k), k=1, 2, \dots\}$ of the fractional phase and frequency process. These τ -domain measures of oscillator stability are uniquely related to the f -domain measure $S_y(f)$ by means of the Mellin transform [1]. Notice that they approximate structure functions of the "observed" noise when the effect of the "system" filter is marginal. The widely accepted measures of oscillator stability in precise timekeeping application—the rms fractional frequency deviation and the Allan variance—are examples from these stability functions.

C. τ -Domain to τ -Domain and τ -Domain to f -Domain Transformations

Specification of oscillator stability in τ -domain and f -domain are interrelated. Given the PSD $S_y(f)$, one can evaluate the structure functions $D_x^{(k)}(\tau)$ and $D_y^{(k)}(\tau)$ via (9) and (12). On the other hand, given $D_y^{(m)}(\tau)$ or $D_x^{(m)}(\tau)$, we can find $S_y(f)$ via the Mellin transform or other techniques discussed in [1]. Furthermore, the structure functions $D_y^{(m)}(\tau)$ for $m \leq 5$ have been shown to be interrelated in a simple way [1]. The same is true for $D_x^{(m)}(\tau)$. In the remainder of this paper, we shall demonstrate these relationships by considering the problem of translation between a rms fractional frequency deviation $\Delta f(\tau)/f_0$ vs τ plot and an Allan variance $\sigma_y(\tau)$ vs τ plot, and obtaining $S_y(f)$ from $\Delta f(\tau)/f_0$.

Under the regular assumptions that $\sigma_y^2 = \int_0^\infty S_y(f) df$ and $\sigma_x^2 \triangleq \int_0^\infty S_y(f)/(2\pi f)^2 df$ are finite, it was shown in [1] that $\sigma_y(\tau)$ and $\Delta f(\tau)/f_0$ are interrelated via

$$\frac{\Delta f(\tau)}{f_0} = \sqrt{\frac{1}{2}\sigma_y^2(\tau) + \frac{1}{2}\sum_{k=1}^{\infty}\sigma_y^2(2^k\tau)} \quad (44)$$

and

$$\sigma_y^2(\tau) = 2\left\{\left[\frac{\Delta f(\tau)}{f_0}\right]^2 - \left[\frac{\Delta f(2\tau)}{f_0}\right]^2\right\} \quad (45)$$

directly in the τ -domain. Let us consider the following typical hydrogen maser taken from [9, Fig. 9] with a $\sigma_y(\tau)$ vs τ plot as depicted in Fig. 8. For $\tau \geq 10^5$ sec, we have added the portion: $\sigma_y^2(\tau) = 3\sigma_x^2\tau^{-2}$. This additional portion of the curve results from considering the limiting behavior of $\sigma_y^2(\tau)$ as $\tau \rightarrow \infty$ [5]. The τ -domain translation using (44) is performed and the result is indicated in Fig. 8. Conversely, if the curve for $\Delta f(\tau)/f_0$ is given, we can simply use (45) to obtain $\sigma_y(\tau)$.

Finally, we shall derive a formula from which one can compute $S_y(f)$ from $D_y^{(1)}(\tau)$. Using [1, eq. V-8] and integrating by parts, we have

$$S_y(f) = \lim_{R \rightarrow \infty} \left[2\pi f D_x^{(1)}(\tau) \sin 2\pi f R - (2\pi R)^2 \int_0^R D_x^{(1)}(\tau) \cos 2\pi f \tau d\tau \right] \quad (46)$$

⁴Stability functions and structure functions are used interchangeably.

Since we can always choose $f \cdot R = k$ an integer, (46) simplifies to

$$S_y(f) = \lim_{k \rightarrow \infty} (2\pi f)^2 \int_0^{k/f} D_x^{(1)}(\tau) \cos 2\pi f \tau d\tau \quad (47)$$

For computational purposes, (47) can be rewritten as

$$S_y(f) = (2\pi f)^2 \lim_{k \rightarrow \infty} \sum_{\ell=0}^{k-1} \int_0^{1/f} D_x^{(1)}\left(\tau + \frac{\ell}{f}\right) \cos 2\pi f \tau d\tau \quad (48)$$

Similar results can be established for $D_y^{(1)}(\tau)$ and $S_x(f)$. Given any phase or frequency structure functions of order 5 or less, we can easily compute $D_x^{(1)}(\tau)$ or $D_y^{(1)}(\tau)$. Then $S_y(f)$ can be evaluated from (48). As an example, we can compute the PSD $S_y(f)$ of the hydrogen maser example using (13) and (48). An alternative way would be to estimate $S_y(f)$ directly from $\sigma_y^2(\tau)$ by the table look-up method described in [5], assuming a sum of power-law component PSD.

REFERENCES

- [1] W. C. Lindsey and C. M. Chie, "Theory of Oscillator Instability Based Upon Structure Functions," Proc. IEEE, Vol. 64, pp. 1652-1666, December, 1976.
- [2] W. C. Lindsey and C. M. Chie, "Interpretation of Frequency and Phase Instability Using Structure Functions," Proc. Precise Time and Time Interval Applications and Planning Meeting, November, 1976.
- [3] J. A. Barnes et al., "Characterization of Frequency Stability," IEEE Trans. Instrum. Meas., vol. IM-20, pp. 105-120, May, 1971.
- [4] E. Hafner, "Outlook for Precision Frequency Control in the 1980's," Proc. Seventh Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, Dec. 2-4, 1975, Goddard Space Flight Center, Greenbelt, Maryland, pp. 101-119.
- [5] W. C. Lindsey and C. M. Chie, "Identification of power-law Type Oscillator Phase Noise Spectra from Measurements," submitted for publication to IEEE Trans. Instrum. Meas.
- [6] D. W. Allan, The Measurement of Frequency Stability of Precision Oscillators, NBS Technical Note 669, CODEN: NBTNAE, U.S. Superintendent of Documents, Washington, D.C., 1975.
- [7] Algie L. Lance, et al., "Automated phase Noise Measurements in the Frequency Domain," in this Proceedings, 1977.
- [8] W. C. Lindsey and C. M. Chie, "Frequency Multiplication Effects on Oscillator Instability," submitted for publication to IEEE Trans. Instrum. Meas.

- [9] V. Reinhardt, "Frequency Stability Requirements for Two Way Range Tracking," Proc. Seventh Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, Dec. 2-4, 1975, Goddard Space Flight Center, Greenbelt, Maryland, pp.265-283.

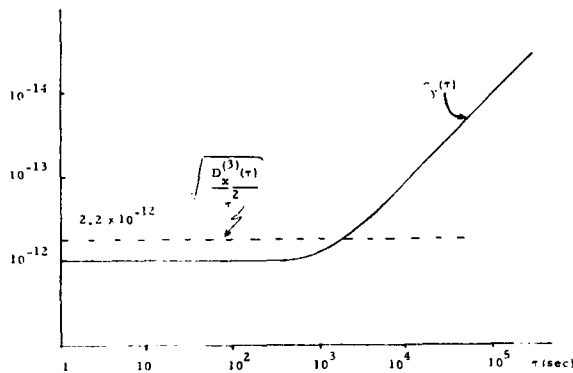


FIG. 2 An Oscillator with both random fluctuations of Flicker Noise FM and a Non-Random linear fractional frequency drift of 10^{-10} per day.

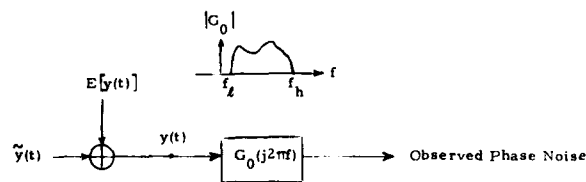


FIG. 3 Model for Observing Phase Noise.

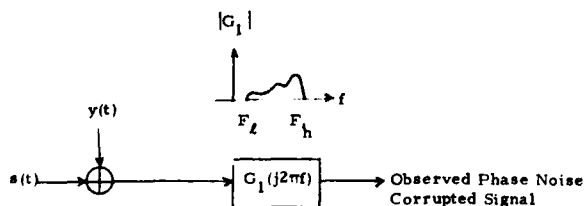


FIG. 4 Model for Assessing Effect of Phase Noise in System Performance.

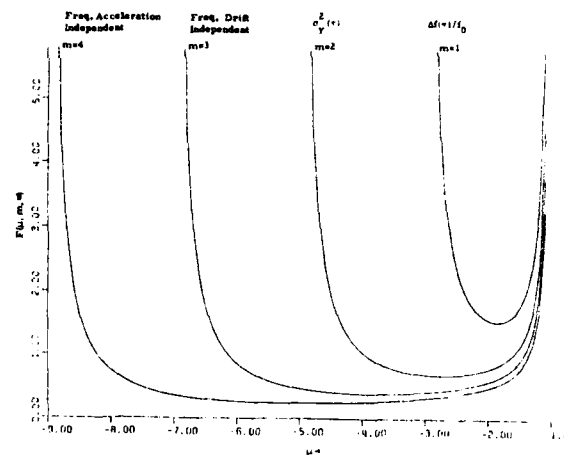
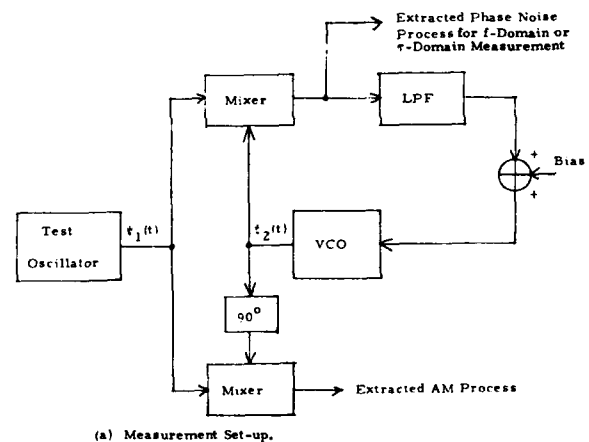
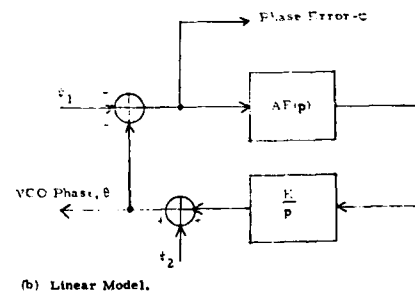


FIG. 1 A Plot of $F(u, m)$ for $m=1,2,3,4$ (Universal curves for identifying power-law type components).



(a) Measurement Set-up.



(b) Linear Model.

FIG. 5 Two-Oscillator Test Set-up and Model.

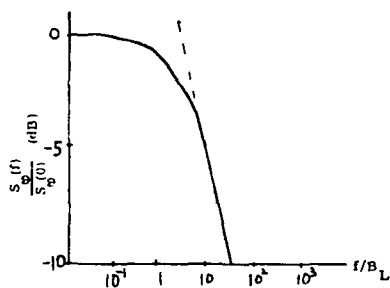
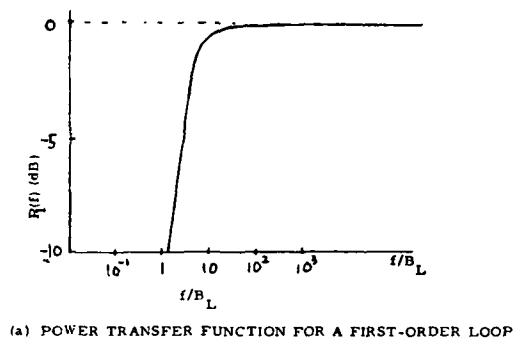


FIG. 6 Power Transfer Function and Observed PSD

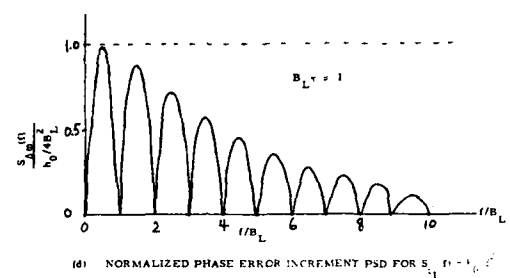
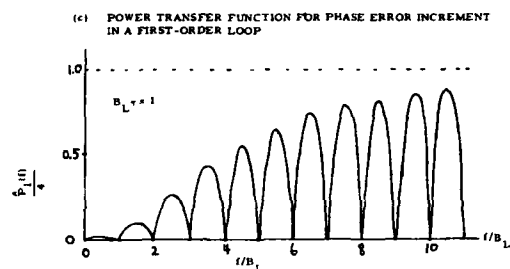


FIG. 6 (Continued)

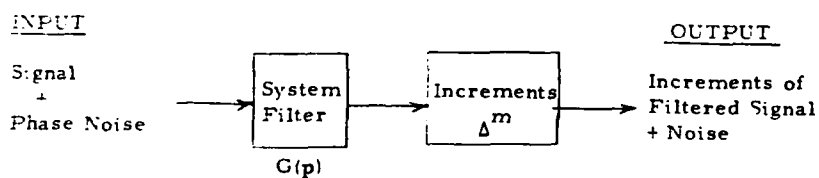


Figure 7. A Block Diagram for Determining Phase Noise Degradation of System Performance.

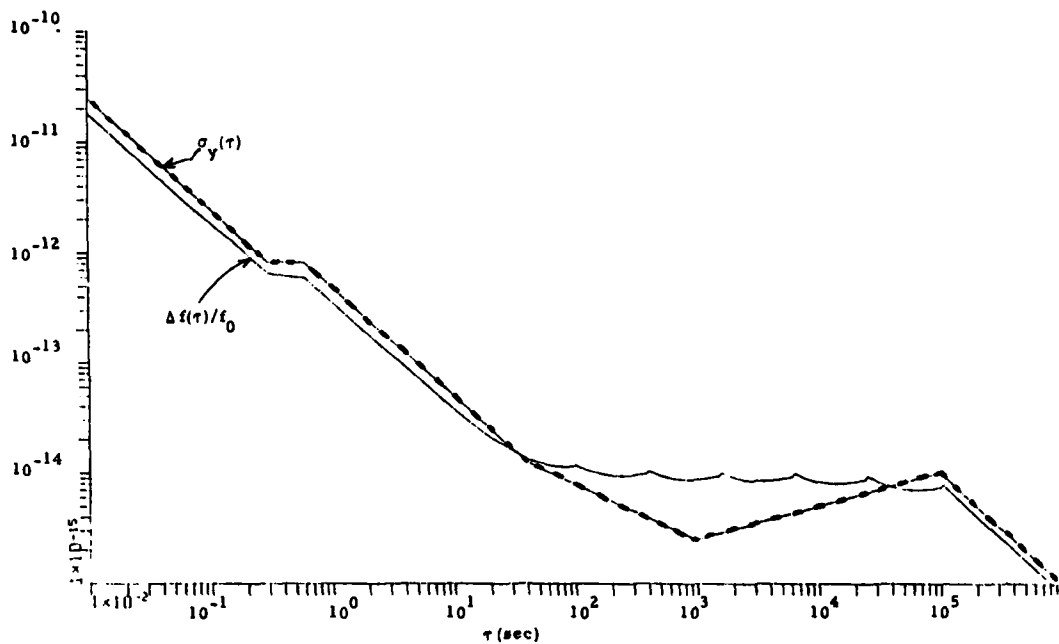


FIG. 8 $\sigma_y(T)$ for Non-automated NASA Hydrogen Maser.

SYSTEM TYPE	INCREMENT (m)	PHASE NOISE SOURCE	SYSTEM FILTER
One-Way Doppler (Doppler Rate)	1 (2)	Transmitter Oscillator Receiver VCO Receiver Reference	$11(s)^*$ $1-H(s)$ 1
Two-Way Doppler (Doppler Rate)	1 (2)	Transmitter Oscillator Vehicle VCO Ground VCO Ground Reference	$H_g(s) H_v(s)$ $H_g(s) [1-H_v(s)]$ $[1-H_g(s)]$
PLL Tracking (Skip Performance)	1	Transmitter Oscillator Local VCO	$1-H(s)$ $1-H(s)$
Communication System (Bit Error Rate)	1	Transmitter Oscillator Local VCO	$1-H(s)$ $1-H(s)$
One-Way Ranging (Range Rate)	0 (1)	Transmitter Oscillator Receiver VCO Receiver Reference	$11(s)$ $1-H(s)$ 1
Two-Way Ranging (Range Rate)	0 (1)	Transmitter Oscillator Vehicle VCO Ground VCO Ground Reference	$H_g(s) H_v(s)$ $H_g(s) [1-H_v(s)]$ $[1-H_g(s)]$ 1

* $H(s)$ - Closed Loop Filter

TABLE I Phase Noise Degradation of System Performance.

ESTIMATION OF THE TWO-SAMPLE VARIANCE WITH A LIMITED NUMBER OF DATA

P. Lesage and C. Audoin
Laboratoire de l'Horloge Atomique
Equipe de Recherche du CNRS
Bâtiment 221 - Université Paris-Sud
91405 - Orsay - France

Summary

The uncertainty in the estimation of the two-sample variance which results from the finite number of available data is considered. It is shown that the finite number of measurements does not introduce bias in the time domain measure of frequency stability. The precision in the estimation of the two-sample variance is specified as a function of the number m of available data. The result is valid for $m \geq 2$. The effect of a dead time between successive measurements is pointed out.

Experimental results confirm that frequency fluctuations are normally distributed for white frequency and white phase noises. Experimental results concerning the dependence of the precision in the estimation of the two-sample variance on the number of data agree with theoretical conclusions.

Introduction

The definition of the time domain measure of frequency stability ¹ assumes an infinite number of measurements, which, obviously, is unrealistic. If it is feasible to accumulate a large number m of samples for small values of τ , this is not yet true for τ larger than say 1 day. Consequently, only an estimate of the average of the two sample variance $\sigma_y^2(\tau)$ can be calculated in practice. The question then arises of the comparison of the estimated average and of the average value of the two sample variance $\sigma_y^2(\tau)$, especially for small values of m ($m \geq 2$).

This problem was considered in a previous work ². It is partly completed here to take into account a possible dead-time between successive measurements, and theoretical predictions of the effect of this dead time are experimentally checked in the case of the white phase noise. Experimental results concerning white frequency noise - in the absence of a dead time - are reported. Furthermore, an error which has been pointed out to us by Dr. Barnes is removed in the present paper.

The estimated two-sample variance

We consider the output voltage of a signal generator. It may be written as :

$$V(t) = V_0 \sin [2\pi\nu_0 t + \varphi(t)] \quad (1)$$

where $\varphi(t)$ represents the instantaneous phase fluctuations of the output signal, V_0 and ν_0 are its nominal amplitude and frequency, respectively. We assume that the amplitude fluctuations of the signal are negligible and that its frequency fluctuations are small. We let ¹ :

$$x(t) = \frac{\varphi(t)}{2\pi\nu_0} \quad (2)$$

The instantaneous fractional frequency deviation $y(t)$ is then :

$$y(t) = \dot{x}(t) \quad (3)$$

and the mean value of $y(t)$ over the time interval $[t_k, t_k + \tau]$ equals :

$$\bar{y}_k = \frac{1}{\tau} \int_{t_k}^{t_k + \tau} y(t) dt = \frac{1}{\tau} [x(t_k + \tau) - x(t_k)] ; t_k = t_0 + kT \quad (4)$$

where T is the time interval between the beginnings of two successive measurements. The quantity y_k is directly accessible to experiment by frequency counting or period measurement ^{1,3}. If τ_k is the measured period of the signal, we have :

$$\bar{y}_k = 1 - \frac{\tau_k}{T} ; \tau = \langle \tau_k \rangle \quad (5)$$

The averaged two-sample variance (A.T.S.V.) of the fractional frequency fluctuations is then defined as ⁴ :

$$\langle \sigma_y^2(2, T, \tau) \rangle = \frac{1}{2} \langle (\bar{y}_{k+1} - \bar{y}_k)^2 \rangle \quad (6)$$

This definition assumes an infinite number of measurements. In fact, only a finite number m of measurements of \bar{y}_k can be performed, and the *estimated* two sample variance (E.T.S.V.), which is a function of m , is calculated in practice as follows :

$$\hat{\sigma}_y^2(2, T, \tau, m) = \frac{1}{2(m-1)} \sum_{k=1}^{m-1} (\bar{y}_{k+1} - \bar{y}_k)^2 \quad (7)$$

This formula shows that the E.T.S.V. is a random function of m . The question then arises of the comparison of the A.T.S.V. and the E.T.S.V. Equation (7) gives :

$$\langle \hat{\sigma}_y^2(2, T, \tau, m) \rangle = \langle \sigma_y^2(2, T, \tau) \rangle \quad (8)$$

This shows that the finite number of measurements does not introduce any bias in the estimation of the two-sample variance.

Variance of the estimated two-sample variance

The E.T.S.V. being a random function of m , we need to characterize the uncertainty in the estimation. We thus introduce the variance of the E.T.S.V., accordingly to the common understanding of a variance. We set :

$$\sigma_y^2[\hat{\sigma}_y^2(2, T, \tau, m)] = \langle [\hat{\sigma}_y^2(2, T, \tau, m) - \langle \sigma_y^2(2, T, \tau) \rangle]^2 \rangle \quad (9)$$

With the expression (7) of the E.T.S.V., we get :

$$\sigma^2[\hat{\sigma}_y^2(2, T, \tau, m)] = \left[\frac{1}{2(m-1)} \right]^2 \sum_{i=1}^{m-1} \sum_{j=1}^{m-1} \beta_i \beta_j > \quad (10)$$

with

$$\beta_i = (\bar{y}_{i+1} - \bar{y}_i)^2 - 2 \langle \sigma_y^2(2, T, \tau) \rangle \quad (11)$$

We consider, at first, the general dependence of $\hat{\sigma}_y^2(2, T, \tau, m)$ on the value of m .

The classical law of large numbers⁵ which states that the true variance of a sum of $(m-1)$ uncorrelated random variables decreases as $1/(m-1)$, even for small values of $(m-1)$, does not apply here. We are considering the quantities β_i which are correlated because two adjacent differences, $y_{i+1} - y_i$ and $y_{i+2} - y_{i+1}$ are obviously not independent.

Equation (10) can also be written as :

$$\sigma^2[\hat{\sigma}_y^2(2, T, \tau, m)] = \frac{1}{m-1} \left[\frac{\Gamma_0}{4} + \frac{1}{2(m-1)} \sum_{k=1}^{m-2} (m-1-k) \Gamma_k \right] \quad (12)$$

with

$$\Gamma_k = \langle \beta_i \beta_{i-k} \rangle \quad (13)$$

Γ_k , which does not depend on m , represents the autocorrelation coefficient of β_i and β_{i-k} . The same data being used in two adjacent pairs, the autocorrelation coefficient Γ_k (and possibly others : see Appendix) differs from zero. Equation (12) then shows that the $1/(m-1)$ dependence also occurs for the considered random variables, but asymptotically for large enough values of $m-1$. We will derive the exact dependence in the following.

For further calculation, we need to assume that the quantities $(y_{i+1} - y_i)$ are normally distributed. It seems reasonable to assume that $y(t)$ and therefore y_k are Gaussian processes because the frequency fluctuations are the result of a large number of independent perturbations. Furthermore, we have experimentally checked that y_k is a Gaussian process in the cases of the white phase, white frequency and flicker frequency noises (see reference 2, and this work). The Gaussian character being retained in a linear transformation, $(y_{i+1} - y_i)$ is also normally distributed. We are then allowed to transform the fourth-order momentum involved in (10) into a sum of second-order momenta, and we obtain :

$$\sigma^2[\hat{\sigma}_y^2(2, T, \tau, m)] = \frac{1}{2(m-1)^2} \times \sum_{i=1}^{m-1} \sum_{j=1}^{m-1} \langle (\bar{y}_{i+1} - \bar{y}_i)(\bar{y}_{j+1} - \bar{y}_j) \rangle^2 \quad (14)$$

Relation between the variance of the estimated two sample variance and the spectral density of frequency fluctuations

We consider the autocorrelation function of $x(t)$ defined by * :

* The autocorrelation function $R_x(\theta)$ is defined for white and flicker noises of phase. For white, flicker and random frequency noises, the phase is not stationary⁶. One should then consider the autocorrelation function of $y(t)$ ⁷. This will not be made here for sake of brevity. Relation (22) is valid for all the usually considered noise processes.

$$R_x(\theta) = \langle x(t) x(t-\theta) \rangle \quad (15)$$

Relation (14) can then be written as :

$$\sigma^2[\hat{\sigma}_y^2(2, T, \tau, m)] = \frac{1}{2(m-1)^2} \sum_{i=1}^{m-1} \sum_{j=1}^{m-1} \left\{ 4 R_x[(i-j)T] - 2 R_x[(i-j)T - \tau] - 2 R_x[(i-j)T + \tau] - 2 R_x[(i-j-1)T] - 2 R_x[(i-j+1)T] + R_x[(i-j-1)T - \tau] + R_x[(i-j-1)T + \tau] + R_x[(i-j+1)T - \tau] + R_x[(i-j+1)T + \tau] \right\} \quad (16)$$

The autocorrelation function $R_x(\theta)$ is connected to the one-sided power spectral density of fractional frequency fluctuations $S_y(f)$ by :

$$R_x(\theta) = \frac{1}{4\pi^2} \int_0^\infty \frac{1}{f^2} S_y(f) \cos 2\pi f \theta df \quad (17)$$

We consider the set of five independent noise processes with the following power law spectral densities :

$$S_y(f) = h_\alpha f^\alpha ; \quad \alpha = 2, 1, 0, -1, -2, \quad f > 0 \quad (18)$$

We take into account the finite bandwidth of the signal source and of the circuits which are used to filter the signal before its processing by the frequency fluctuation measuring system. Two different expressions of the system function $H(f)$ of this filter are generally used 1,2,7 :

$$H_1(f) = \begin{cases} 1 & ; 0 \leq f \leq f_c \\ 0 & \text{elsewhere} \end{cases} \quad (19)$$

$$H_2(f) = \frac{1}{1 + j \frac{f}{f_c}} \quad (20)$$

The spectral density of the observed fractional frequency fluctuations is then :

$$S_y(f) = h_\alpha f^\alpha |H(f)|^2 \quad (21)$$

Introducing (17) into (16) we obtain the following expression of the variance of the E.T.S.V. :

$$(m-1)^2 \sigma^2[\hat{\sigma}_y^2(2, T, \tau, m)] = 8(m-1) \left\{ \int_0^\infty S_y(f) \frac{\sin^2 \pi \tau T \sin^2 \pi f T}{(\pi f T)^2} df \right\}^2 + 16 \sum_{k=1}^{m-2} (m-k-1) \left\{ \int_0^\infty S_y(f) \cos 2k\pi f T \frac{\sin^2 \pi \tau T}{(\pi f T)^2} df \right\}^2 \quad (22)$$

The integrals involved in (22) converge for $f = 0$ and $f \rightarrow \infty$, for the filtered spectral densities considered here. The variance of the E.T.S.V. is therefore defined without any approximation.

Relative uncertainty on the characterization of frequency stability

Relative uncertainty on estimated two-sample variance

We introduce Δ the deviation of $\hat{\sigma}_y^2(2, T, \tau, m)$ relative to $\langle \sigma_y^2(2, T, \tau) \rangle$

$$\Delta = \frac{\hat{\sigma}_y^2(2, T, \tau, m) - \langle \sigma_y^2(2, T, \tau) \rangle}{\langle \sigma_y^2(2, T, \tau) \rangle} \quad (23)$$

The quantity Δ is a random function of m . The standard deviation $\sigma(\Delta)$ defines the relative uncertainty in the determination of $\langle \sigma_y^2(2, T, \tau) \rangle$ due to the finite number of measurements. From relation (23) we obtain

$$\sigma(\Delta) = \frac{1}{\langle \sigma_y^2(2, T, \tau) \rangle} \left\{ \sigma_y^2 [A_y^2(2, T, \tau, m)] \right\}^{1/2} \quad (24)$$

The dependence of $\sigma(\Delta)$ on the available number of measurements can then be derived from expressions (22) and (24). It occurs that the final result is the same for the two considered system functions $H_1(f)$ and $H_2(f)$ in the case $T = \tau$, and for $\alpha = -2$ to 2 . In the presence of a dead time, the same is true for white noise of phase. Table 1 shows the expressions of $\sigma(\Delta)$, as a function of m , for the considered noise processes, and for $T = \tau$. For $\alpha = 2, 1, 0, -2$, the results are valid for $m \geq 2$. In the case $\alpha = -1$, the expression given in table 1 is valid within a precision of 1 % for $m \geq 5$; for smaller values of m , the exact results are specified in Table 2.

It occurs that the uncertainty in the estimation of the two sample variance depends on the dead-time $T - \tau$ between successive measurements. Up to now, we have only derived a result useful for the interpretation of an experiment reported below. More detailed results will be published later. For white noise of phase, we have

$$\sigma(\Delta) = \frac{\sqrt{12m-16}}{2(m-1)} \quad (25)$$

This result is valid for $2\pi f_c T \gg 1$ and $2\pi f_c (T - \tau) \gg 1$.

Precision of the time domain measure of frequency stability

In practice, the frequency stability of a frequency source is characterized by the square root of the experimentally determined F.T.S.V., i.e., with $[A_y^2(2, T, \tau, m)]^{1/2}$. Therefore, we also introduce δ given by:

$$\delta = \frac{[A_y^2(2, T, \tau, m)]^{1/2} \cdot \langle \sigma_y^2(2, T, \tau) \rangle^{1/2}}{\langle \sigma_y^2(2, T, \tau) \rangle^{1/2}} \quad (26)$$

$\sigma(\delta)$ specifies the precision of the time domain measure of frequency stability, as a function of the number of available data.

We will now assume that the condition $\delta \ll 1$ is fulfilled i.e. that the number of measurements is large enough to get a good confidence on the frequency stability measure. From equations (26) and (24), we then obtain the following expression:

$$\sigma(\delta) = \frac{1}{2} \sigma(\Delta) \quad \text{for } \delta \ll 1 \quad (27)$$

From Table 1 and relation (27), we then obtain the following result for $T = \tau$, valid within a precision of 10 % for $m > 5$

$$\sigma(\delta) = K_\alpha m^{-1/2} \quad (28)$$

with

$$\begin{aligned} K_2 &= K_1 = 0.99 \\ K_0 &= 0.87 \\ K_{-1} &= 0.77 \\ K_{-2} &= 0.75 \end{aligned} \quad (29)$$

In the presence of a dead-time, such as $2\pi f_c (T - \tau) \gg 1$ and $2\pi f_c \tau \gg 1$, we obtain $\sigma(\delta) = 0.87 m^{-1/2}$ for the white noise of phase case, which differs significantly from the result $\sigma(\delta) = 0.99 m^{-1/2}$ established within the condition $T = \tau$.

Experimental results

Some experimental results have been published previously². They are completed, using a cesium beam frequency standard and hydrogen masers to experience the cases $\alpha = 0$ and 2 .

White frequency noise

Appropriate signals are frequency synthesized from the 5 MHz output of a cesium beam frequency standard, and drive a triple heterodyne receiver which is connected to the 1.42 GHz output of a hydrogen maser. A beat note at the frequency of 5.75 kHz is obtained. It is filtered in a selective amplifier, with a bandwidth of 40 Hz. The duration of 57 500 periods of the beat note is measured. Therefore τ equals 10 s. The dead time amounts to 0.17 ms and is assumed to be negligible.

For $\tau = 10$ s, the cesium beam frequency standard shows a frequency stability $\sigma_y = 1.7 \times 10^{-11}$ and we are well within the range of τ values for which $\sigma(\tau)$ varies as $\tau^{-1/2}$. The frequency generator is then perturbed by white noise of frequency.

A total number of 7 500 measurements has been performed. Counting results were recorded on a magnetic tape for the purpose of computer analysis. Figure 1 shows that counting times τ_k are normally distributed. Therefore, y_k and $y(t)$ are also normally distributed.

Afterwards, the measurements have been separated into 7 500/m sets of m measurements with $2 \leq m \leq 800$. The E.T.S.V. has been computed for each set of given m , and the relative uncertainty $\frac{1}{2} \sigma(\delta)$ i.e. $\sigma(\delta)$ for m large enough has been determined from the 7 500/m values of the E.T.S.V. The variation of $\frac{1}{2} \sigma(\delta)$ are depicted on figure 2. The experimental result is

$$\sigma(\delta) = 0.84 m^{-0.49} \quad \text{for } \alpha = 0 \text{ and } T = \tau \quad (30)$$

This agrees satisfactorily well with the theoretical result.

White phase noise

A beat note is obtained between two hydrogen masers, at a frequency of 2 Hz. It is low-pass filtered, with a bandwidth of 6 Hz. The duration of 10 periods of the beat note is measured. We then have $\tau = 5$ s. The dead-time equals 0.5 s and the condition $2\pi f_c (T - \tau) \gg 1$ is satisfied. In this experiment, H-masers are perturbed by white noise of phase, and we have $\sigma_y = 9 \times 10^{-14}$ for $\tau = 5$ s.

A total number of 7 500 measurements has been processed as above. Figure 3 shows that counting results are normally distributed and that $y(t)$ represents a Gaussian process. The experimentally determined variation of $\frac{1}{2} \sigma(\delta)$ (i.e. $\sigma(\delta)$ for m large enough) are shown on figure 4. The experimental result is

$$\sigma(\delta) = 0.83 m^{-0.47} \quad (31)$$

for $\alpha = 2$ and $2\pi f_c \tau \gg 1$, $2\pi f_c (T - \tau) \gg 1$, in good agreement with theory, in the presence of a dead-time.

Conclusion

We have shown that the finite number of data does not introduce bias in the time domain measure of frequency stability. We have compared the variance of the average value of the two sample variance with the classical result of the law of large numbers, and we have shown the reasons of the small discrepancy between the relevant theoretical results, for small values of m .

The variation of the variance of the average value of the two sample variance, and of its square

root, as a function of the number of available data, is given. The precision in the time domain measure of frequency stability can then be stated, accordingly to the number of data used in its estimation. The effect of dead-time between successive measurements is pointed out.

Experimental results concerning white phase and white frequency noises are reported. The frequency fluctuations are normally distributed and theoretical results are experimentally verified for the two aforementioned kinds of noise.

Appendix

The calculation of Γ_k for a normally distributed noise process shows the results summarized below, where the Γ_k which differs from zero are specified.

$$\begin{aligned} \alpha = 2 & \quad \Gamma_0, \Gamma_1 \\ \alpha = 1 & \quad \Gamma_0, \Gamma_1, \Gamma_2 \\ \alpha = 0 & \quad \Gamma_0, \Gamma_1 \\ \alpha = -1 & \quad \Gamma_k \\ \alpha = -2 & \quad \Gamma_0, \Gamma_1 \end{aligned}$$

References

- (1) J.A. Barnes, A.R. Chi, L.S. Cutler, D.J. Healey, D.B. Leeson, T.E. McGunigal, J.A. Mullen, W.L. Smith, R.L. Sydnor, R.F.C. Vessot, G.M.R. Winkler, "Characterization of Frequency Stability". IEEE IM-20, n° 2 pp. 105-120, May 1971
- (2) P. Lesage and C. Audoin, "Characterization of Frequency Stability : Uncertainty due to the finite number of Measurements". IEEE IM-22 n° 2, pp. 157-161, June 1973 ; IEEE IM-24, p. 86, March 1975 ; IEEE IM-25, n° 3, p. 270, September 1976
- (3) L.S. Cutler, C.L. Searle, "Some aspects of the theory and measurement of frequency fluctuations in frequency standards". Proc. IEEE Vol. 54, pp. 136-154, Feb. 1966
- (4) D.W. Allan, "Statistics of atomic frequency standards". Proc. IEEE, Vol. 54, pp. 221-230, Feb. 1966
- (5) A. Papoulis, "Probability, random variables and stochastic processes". Mac Graw Hill, 1965, pp. 263-266
- (6) E. Boileau, B. Picinbono, "Statistical study of phase fluctuations and oscillator stability". IEEE IM-25, n° 1, pp. 66-75, March 1976
- (7) P. Lesage, "Caractérisation de la stabilité de fréquence d'oscillateurs dans le domaine temps et dans le domaine fréquence". Thèse de Doctorat de 3e cycle, Orsay 1973

TABLE 1

Summary of theoretical results for $\tau = \tau_c$

Spectral density of fractional frequency fluctuations	$S_y(f)$	$h_2 f^2$	$h_1 f$	h_0	$h_{-1} f^{-1}$	$h_{-2} f^{-2}$
Relative uncertainty on the estimated two sample variance	$\sigma(\Delta)$	$\frac{\sqrt{35m-53}}{3(m-1)}$ $2\pi f \tau \gg 1$	$\frac{\sqrt{35m-53}}{3(m-1)}$ $2\pi f \tau \gg 1$	$\frac{\sqrt{3m-4}}{(m-1)}$ $2\pi f \tau \gg 1$	$\frac{\sqrt{2.3m-2.6}}{(m-1)}$ $m > 5$	$\frac{\sqrt{9m-10}}{2(m-1)}$

TABLE 2

Value of the relative uncertainty on the estimated two-sample variance for flicker noise of frequency, for small values of m.

m	2	3	4	5
$\sigma(\Delta)$	1.41	1.02	0.85	0.74

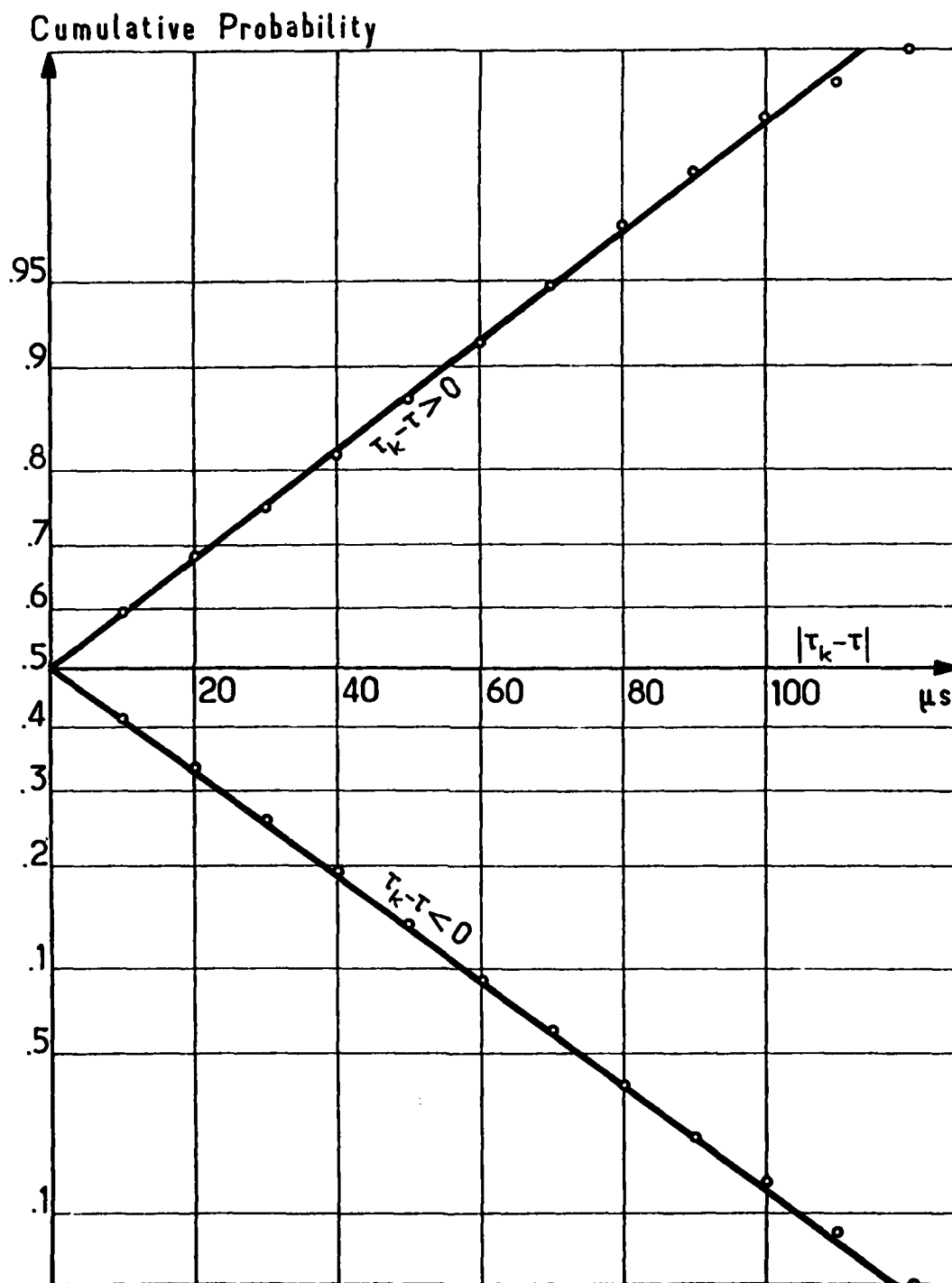


Fig. 1 : Distribution of counting times for white frequency noise ($\alpha = 0$, $\tau = 10$ s) in Galtonian coordinates. Circles represent the cumulative probability corresponding to $|\tau_k - \tau|$, with $\tau = \langle \tau_k \rangle$. Solid line corresponds to the normal distribution of same width.

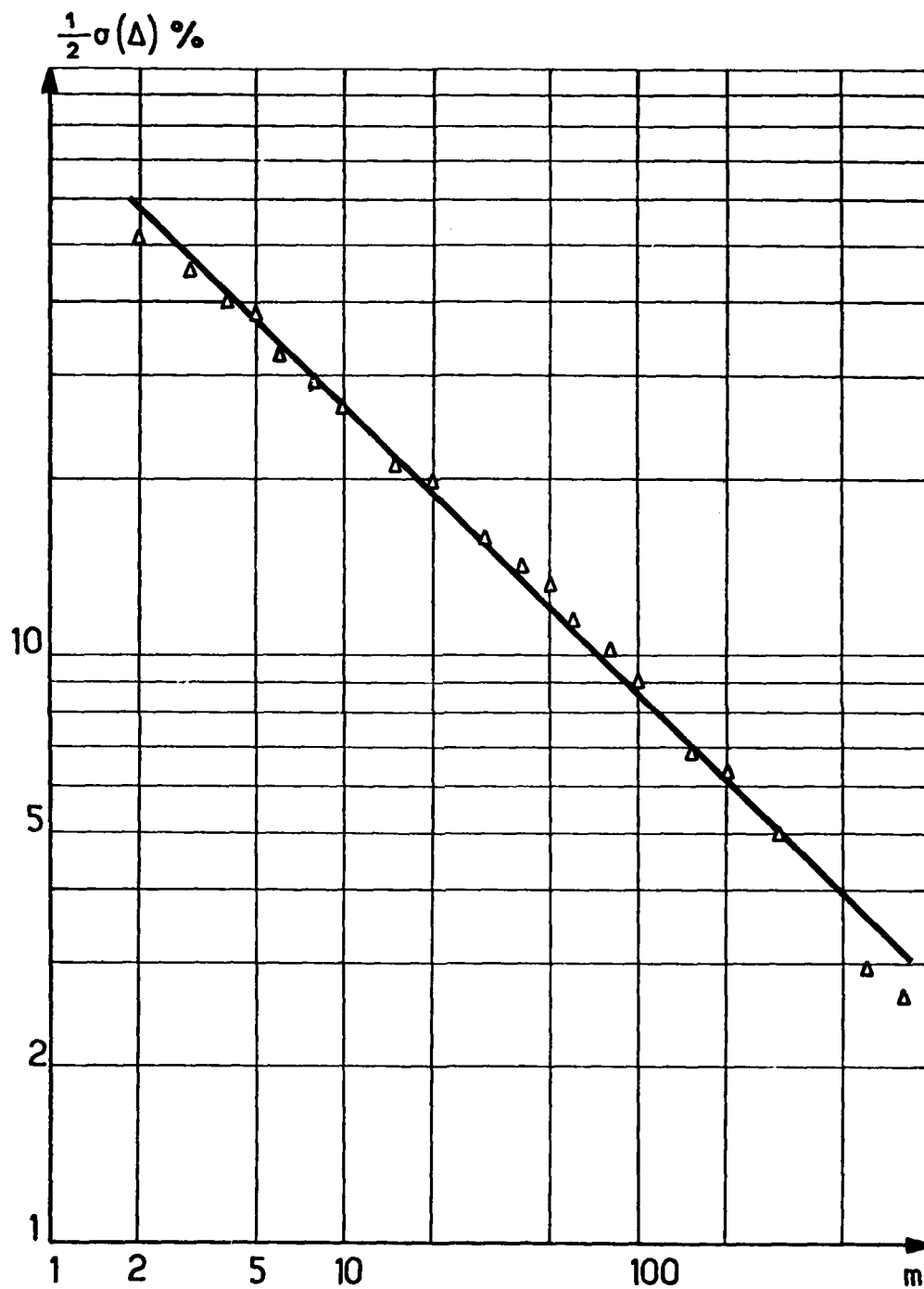


Fig. 2 : Experimental dependence of $\frac{1}{2}\sigma(\Delta)$ on number m of measurements for white frequency noise ($\alpha = 0$, $\tau = 10^{-5}$ s). Circles are experimental results. Solid line is adjusted by method of least squares. Its slope is -0.49 and experimental value of K_0 equals 0.84 .

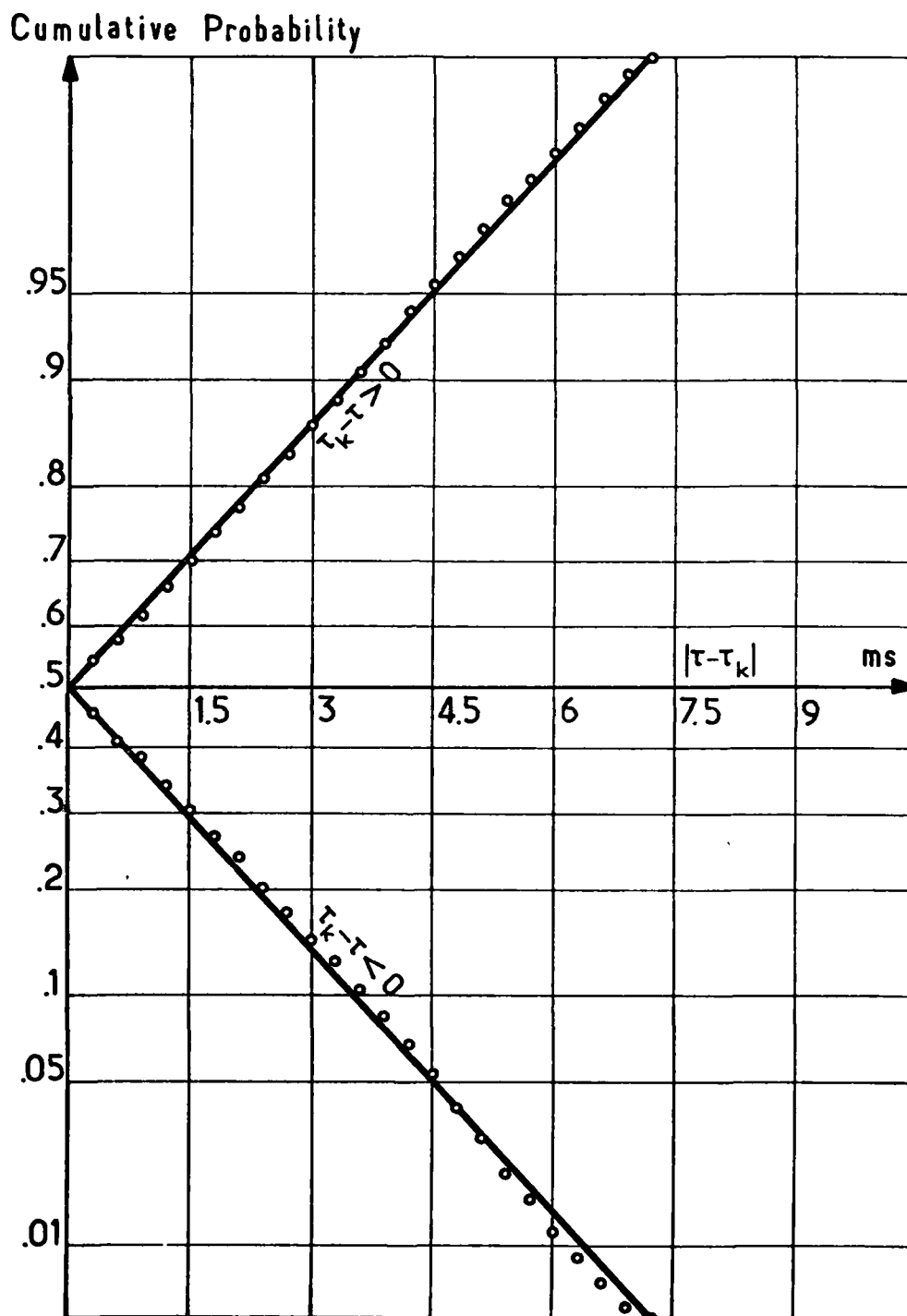


Fig. 3 : Distribution of counting times for white phase noise ($\alpha = 2$, $\tau = 5$ s). Circles represent the cumulative probability corresponding to $|\tau - \tau_k|$, with $\tau = \langle \tau_k \rangle$. Solid line corresponds to the normal distribution of same width.

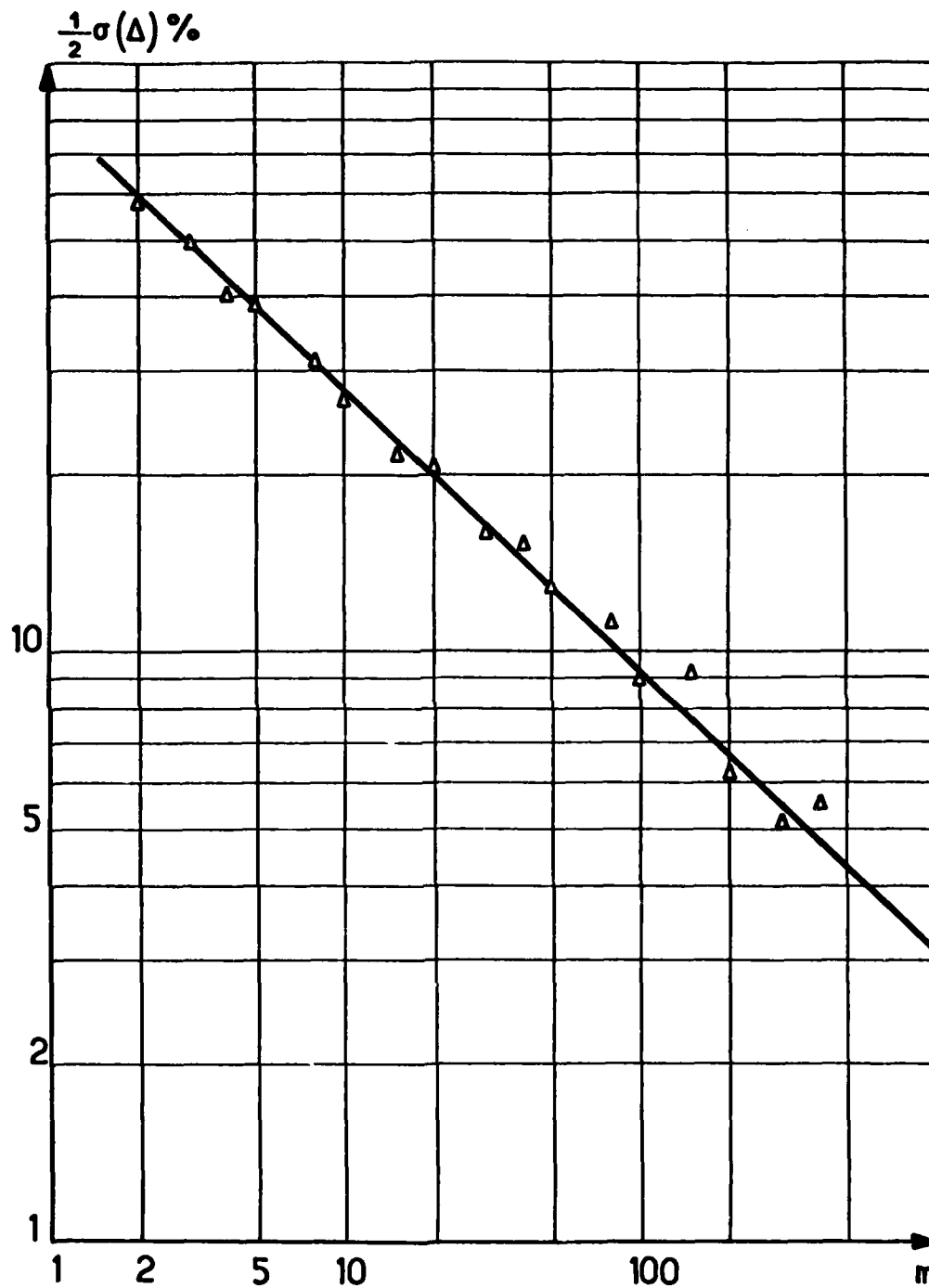


Fig. 4 : Experimental dependence of $\frac{1}{2} \sigma(\Delta)$ on number m of measurements for white phase noise ($\alpha = 2$, $\tau = 5$ s). Circles are experimental results. Solid line is adjusted by method of least squares. Its slope is -0.47 and experimental value of K_0 equals 0.83 .

PREDICTION ERROR ANALYSIS OF ATOMIC FREQUENCY STANDARDS

D. B. Percival
U. S. Naval Observatory
Washington, D. C. 20390

Summary

Predictions of fractional frequency for lead times from 1 to 64 days are made for commercial cesium beam frequency standards compared against A.1 (USNO, MEAN). Two different types of prediction algorithms are studied empirically based upon ARIMA models and fixed memory polynomial filters. Analysis of prediction errors shows that a useful measure of fractional frequency instability for various sampling times may be based upon the empirical cumulative distribution function of prediction errors for various lead times. In particular, the 95% and 50% sample quantiles of prediction errors have several desirable properties as instability measures. First, the problem of defining what is a systematic component in fractional frequency deviates is simplified: if there is no reasonable procedure for dealing *a priori* with a suggested systematic component in a prediction algorithm, then that component will appear as part of the prediction errors. Second, the sample quantiles of the prediction errors may be transformed easily into probabilistic time errors. Third, although a prediction algorithm may be quite complex, quantiles as a measure of fractional frequency instability are easy to communicate to those less familiar with statistical analysis. Fourth, standard statistical procedures exist for estimating confidence intervals for the resulting empirical cumulative distribution function and sample quantiles.

Introduction

The problem of how best to predict future values of the time and frequency of an atomic clock has received much attention in the past ten years.^{1,2,3} The prediction problem is one of practical importance in such areas as utilization of portable clock data, control of time and frequency at remote stations, and time scale formation. This report will first summarize the results of an empirical study of two classes of prediction methods. An analysis of the resulting prediction errors (in cases where these errors are close to white noise) will then show how prediction errors may be used as a simple measure of frequency time instability in a frequency standard. In fact, the use of prediction errors to characterize frequency instability has several important conceptual and analytical advantages.

Description of Data

The data utilized in this study consists of from 3 to 4.5 years of time comparisons of ten commercial cesium beam frequency standards compared against the local, independent time scale generated at the Naval Observatory, denoted as A.1 (USNO, MEAN). Table 1 lists the name of each of the ten frequency standards, the contiguous amount of data available for each standard, and, where applicable, an estimate of a linear frequency drift term made by a linear least squares fit through the fractional frequency values. Figure 1 shows a plot of the relative one-day fractional frequency values for one of the 10 frequency standards, Cs 571. The plot indicates a slow, but pronounced linear drift in frequency. Of the ten clocks studied, five had a significant linear drift in frequency; two others had a drift of a higher order than linear; and three had no significant drift. For

drifts of the magnitude apparent in commercial cesium beam frequency standards, one can see from Figure 1 that long data samples are necessary in order to estimate the drift term accurately.

Prediction Methods

Background

The prediction methods that were examined fall into two general classes: autoregressive integrated moving average (ARIMA) methods and fixed filter methods. Box and Jenkins have given a detailed exposition on ARIMA models.⁴ Barnes has discussed briefly the application of ARIMA models to time and frequency data.² Morrison has an extensive discussion on the fixed filter approach.⁵ A distinctive feature of both classes of prediction methods is that the predictions are based only upon analysis of empirical samples of time series. In particular, no assumptions need be made about the underlying form of the power spectrum of the fractional frequency process.

ARIMA Methods

Following the analysis procedure described by Box and Jenkins, models of the form

$$w_t - \sum_{i=1}^p \phi_i w_{t-i} = \theta_0 + a_t - \sum_{i=1}^q \theta_i a_{t-i}$$

were fit to each of the ten series of one-day fractional frequency values, where w_t is the d -th finite difference at day t of the one-day fractional frequency values, y_t ; $\phi_1, \phi_2, \dots, \phi_p$ are p autoregressive parameters; θ_0 is a parameter which accounts for the slow linear drift in frequency seen in some cesium beam frequency standards; $\theta_1, \theta_2, \dots, \theta_q$ are q moving average parameters; and $a_t, a_{t-1}, a_{t-2}, \dots$ is a sequence of uncorrelated random variables with zero mean and constant variance, i.e., $E(a_t) = 0$, $E(a_t a_{t-1}) = 0$ for $t \neq 1$, and $E(a_t^2) = \sigma_a^2$. Models of the above form are called ARIMA models of order (p, d, q) . Once p , d , and q have been specified, there are $p+q+2$ parameters to be estimated.

By examining the estimated autocorrelation functions and the estimated partial autocorrelation functions for $\{w_t\}$ formed by letting $d = 0, 1$, and 2 , it was found for all ten clocks that these functions damped out rapidly when $d = 1$ (i.e., the first finite difference of the fractional frequency values appeared to be stationary) and that a model of order $(0, 1, 1)$ might be appropriate. (This model is also called an exponentially weighted moving average.) Least squares estimates of the three parameters in the $(0, 1, 1)$ model ($\sigma_a^2, \theta_0, \theta_1$) were obtained. The estimated models were used to predict sequentially the observed series of fractional frequencies. The sequences of one-day prediction errors were then examined using two tests recommended by Box and Jenkins, a portmanteau test on the empirical autocorrelation function of the prediction errors and a cumulative periodogram (white noise) test. For three of the ten clocks, the $(0, 1, 1)$ model was not rejected as being inadequate at the .05 level of significance by either test. For the remaining seven clocks, the next two possible higher order models were then fit,

the (0,1,2) model and the (1,1,1) model. Of the seven remaining clocks, five were not found to be inadequately modeled by one of these two higher order models using both tests at the .05 level of significance. For the two clocks for which none of the three ARIMA models considered was adequate, the (0,1,1) models were used for subsequent comparisons with other prediction methods. (For one clock, Cs 837, a plot of the one-day prediction errors indicated that the lack of fit was due to a gradual increase in the variance of the prediction errors during the last 400 days of data. The other inadequately modeled clock, Cs 591, had an unusual large number of isolated, large prediction errors. It is interesting to note that the two-day average fractional frequency values of Cs 591 were adequately modeled by an ARIMA (0,1,1) model.) Table 2 lists the estimated parameters and the critical values of the two tests for the models of interest for each clock.

For predictions made for k days in advance, where $k > 1$, two procedures were considered. First, the sequence of k day average fractional frequency values was formed for each clock. An ARIMA model was built for each sequence. With these k day models, predictions for k days in advance were made by forecasting one interval (k days) in advance. The second procedure simply consisted of using the one-day ARIMA model to predict k intervals (k days) in advance. An empirical comparison indicated that, while there was usually little practical difference between these two procedures, the second procedure was always as good as or slightly better than the first. The one-day ARIMA models were thus used for all predictions. The prediction equations for the (0,1,1), (0,1,2) and (1,1,1) models are given in appendix 1.

The fractional frequency prediction error made at day t for a lead time of k days is defined here as

$$e_f(k,t) = \frac{1}{k} \sum_{i=1}^k (y_{t+i} - \hat{y}_t(i))/k,$$

where y_t is the observed fractional frequency value at day t ; $\hat{y}_t(i)$ is the predicted fractional frequency value for day $t+i$ made at day t ; $t = 0, k, 2k, \dots, (n-1)k$ to eliminate correlations in the prediction errors due to overlapping prediction intervals; and n is an integer which satisfies the inequality

$$(N/k) - 1 < n \leq N/k,$$

where N is the total number of observed fractional frequency values. This definition of fractional frequency prediction error allows one to translate easily to a phase or time prediction error, $e_p(k,t)$, by

$$e_p(k,t) = 86400k e_f(k,t).$$

The rms fractional frequency error, $e_f(k)$, for a lead time of k days is defined here as

$$e_f(k) = \sqrt{\frac{1}{n} \sum_{j=0}^{n-1} e_f^2(k,jk)/n}.$$

Table 3 lists the rms fractional frequency prediction errors for all ARIMA models of interest for lead times up to 64 days. In order to show the effect of uncertainty in the determination of θ_1 , the linear frequency drift compensation parameter, clocks with a significant frequency drift have been modeled both with a θ_0 term estimated by least squares and with θ_0 set to 0. Generally, θ_0 becomes important for predictions with lead times of 32 days or more. Note also that use of the (0,1,1) model in cases where there is a better fitting (1,1,1) or (0,1,2) model increases the rms error usually by less than 1×10^{-14} . To realize this small decrease in rms prediction error with either the (1,1,1) or (0,1,2) model, one would

need to have a large amount of data to accurately estimate the parameters in these more complex models. For amounts of data less than one or two years, one would have to use the simpler (0,1,1) model.

To gain insight into the effect of a limited amount of data on the determination of the parameters in an ARIMA model, the fractional frequency data for each clock were broken up into non-overlapping segments of 64 days length each. For each segment of 64 days, a simple (0,1,1) ARIMA model was fit. For the case of Cs 571, there were 24 non-overlapping 64 day segments. The θ_1 parameter in the (0,1,1) model ranged from .35 to .95. The resulting rms prediction errors for 1 and 32 day lead times based upon the entire 1586 days worth of data are plotted in Figure 2 as a function of θ_1 . One can see that the rms prediction error is fairly insensitive to variations in θ_1 ; uncertainty in the model parameter θ_1 after a calibration period of 64 days translates at most into a 10% increase in rms prediction error for prediction times of both 1 day and 32 days. Similar results were obtained for the other 9 clocks.

A study of Table 3 and Figure 2 indicates the effect of uncertainty on the rms prediction error of Cs 571. After a 64 day calibration period, one would expect to realize the rms prediction errors shown in Figure 2 by using the (0,1,1) model with $\theta_0 = 0$. After about a year one would have a fairly good estimate of both θ_0 and θ_1 . The realizable rms prediction errors would be close to the errors listed in Table 3 for the (0,1,1) model with $\theta_0 = .003$. Finally, after two or three years, the realizable prediction errors should be close to those for the (1,1,1) model with $\theta_0 = .003$.

Fixed Filter Methods

With the fixed filter approach one fits a polynomial through the r most recent data points and uses the resulting regression equation to forecast future values. Only the r most recent data points are involved in the prediction process. A fixed filter should be well suited to predict a process in which sudden changes occur from time to time. It can be shown that the characteristic square root of Allan variance versus sampling time curve for commercial cesium beam clocks can be simulated well by a suitable combination of white frequency noise plus randomly distributed steps in frequency.^{2,6,7} If this proposed model were true, then a fixed filter predictor should be close to ideal.

Two forms of the fixed filter method were examined. In the order 1 model of length r , the r most recent one-day fractional frequency values are simply averaged to form the prediction for a lead time k . A special case of this method occurs when $r = k$. Then the k -day in advance prediction is simply the most recent k -day fractional frequency value. This procedure has been called the end point prediction method.

For the order 2 model of length r , a linear least squares fit is made through the r most recent one-day fractional frequency values. The resulting regression equation is then used to forecast future values.

For both the order 1 and order 2 models, the filter length r was determined by examining the rms prediction error as r was varied up to a maximum of 256 days. The best r for each lead time k was defined as that r which minimized the k -day rms prediction error. The rms prediction errors and the best r for all ten clocks are listed in Table 4 for the

order 1 model and in Table 5 for the order 2 model.

Comparison of Prediction Methods

A comparison of Tables 3, 4, and 5 indicates that the rms prediction errors for the fixed filter methods were generally somewhat larger than for the ARIMA models. The most notable exceptions were Cs 549 and Cs 660. For longer sampling times, the order 1 fixed filters were better for these two clocks. The order 1 fixed filter was generally superior to the order 2 fixed filter except in the case of Cs 591 and in cases where frequency drift was important. The rms prediction errors for the order 1 fixed filter method were generally quite close to the ARIMA model prediction errors. The fixed filter methods, however, have an advantage over ARIMA methods in the simplicity of their implementation.

Prediction Errors as Instability Measures

The characterization of long term instability of frequency standards has been accomplished in practice in several steps. One obtains a sequence $(x(t), t = 0, \tau, 2\tau, \dots, N\tau)$ of $N+1$ phase measurements between two frequency standards (or a frequency standard and a time scale), where τ is the time between successive measurements. One forms the sequence $(y(\tau) = (x(t) - x(t-\tau))/\tau, t = \tau, 2\tau, \dots, N\tau)$ of N average fractional frequency deviations. A model of the form $y(t) = c(t) + n(t)$ is assumed, where $c(t)$ is some deterministic function of time known as the systematics and $n(t)$ is the noise.⁸ In order to characterize $n(t)$ statistically, $c(t)$ must be removed, and $n(t)$ must be filtered in some fashion to produce an instability measure that converges asymptotically to some value. The variance of the residual sequence is then used for comparison of frequency standards or for further analysis; e.g., a spectral density function is estimated.

This approach to instability characterization has proven quite fruitful. There are two criticisms of this procedure. First, the removal of $c(t)$ is somewhat subjective, and the effect of uncertainty in its removal is not quantified. Second, the quantity to which an instability measure converges may not be immediately useful to all system designers. In particular, some system requirements are stated in terms of tolerable time or phase error. In cases where these criticisms are valid, one can adopt a different approach by equating instability in a frequency standard with a measure of the unpredictability of its fractional frequency. One could, for example, quote rms fractional frequency prediction errors, as done in Table 3. An alternative procedure is to study the empirical cumulative distribution function of the absolute value of the prediction errors for different lead times of interest. This function is shown in Figures 3 and 4 for Cs 571 for 1 day and 32 day prediction times, respectively. The prediction errors referred to in these figures are the errors from the (1,1,1) ARIMA model with a term accounting for linear frequency drift. Since the prediction errors are white and normally distributed approximately, one may assume that they are approximately independent. With this assumption, one can produce confidence intervals for various sample quantiles of interest from the empirical cumulative distribution function using the binomial distribution.⁹ In Figures 3 and 4, separate 95% confidence intervals are shown for the 50% and 95% quantiles. The sample quantiles allow one to make valid probabilistic statements; for example, based upon the past history of Cs 571, one can expect that 95% of the absolute 32 day fractional frequency prediction errors will be less than 8×10^{-14} , or, in terms of phase or

time error, that 95% of the absolute 32 day time prediction errors will be less than 220 nanoseconds.

The difference between instability measures based upon prediction errors and the most commonly used time domain instability measure, the Allan variance, $\sigma_y(\tau)$, is not trivial. Table 6 shows $\sigma_y(\tau)$ and the ratio of $\sqrt{2}\sigma_y(\tau)$ to the rms fractional frequency prediction error of the most sophisticated ARIMA model for each of the ten clocks as a function of lead or sampling time. (For those clocks with a significant linear frequency drift, $\sigma_y(\tau)$ was computed after removal of the frequency drift by linear least squares solution to the fractional frequency values.) Note that this ratio is generally not constant as a function of lead or sampling time. Care must be taken in using $\sigma_y(\tau)$ as a measure of predictability of fractional frequency or time.

The chief disadvantage of instability measures based upon prediction errors is their computational complexity compared to, for example, the Allan variance. There are, however, a number of advantages in considering prediction errors as instability measures. First, unlike other measures of frequency instability, empirical quantiles of the absolute prediction errors are direct probabilistic measures of both time and fractional frequency errors. For system requirements stated in terms of tolerable time or phase error over a certain period of time, a study of empirical distribution errors would indicate how often one would expect a frequency standard to be within tolerance. Second, the quantiles of prediction errors as measures of frequency instability are easy to communicate to those less familiar with statistical analysis. The quantities to which other instability measures converge are often difficult to explain in an intuitive way. Third, the effects of a limited amount of data on the estimation of model parameters and on the elimination of systematic components such as linear frequency drift may be easily quantified. Fourth, standard statistical procedures exist for estimating confidence intervals for the resulting empirical cumulative distribution function and sample quantiles.

Barnes has pointed out that ARIMA models in particular have a number of additional benefits.² After an ARIMA model has been built for a frequency standard, one can easily produce an estimate of the power spectrum of the fractional frequency process, and one knows immediately how to simulate sequences of fractional frequency values with statistical content similar to the original clock data. One can also produce a theoretical Allan variance versus sampling time curve from the parameters in an ARIMA model. In particular, for the (0,1,2) model (of which the (0,1,1) model is a special case), there exists the exact relationship

$$\sigma_y^2(\tau) = \sigma_a^2(2\tau(1+\theta_1^2+\theta_2^2+2\theta_1\theta_2-2\theta_1-2\theta_2)+\tau^{-1}(1+\theta_1^2+\theta_2^2-4\theta_1\theta_2+4\theta_1+22\theta_2)-\tau^{-2}18\theta_2)/6,$$

for $\tau \geq 1$ day. (In fact, for any model of the form (0,1,q), there exists a closed form theoretical Allan variance, although for large q it would be very tedious to derive.) For the (1,1,1) model, an approximate theoretical Allan variance curve can be derived by converting the (1,1,1) model into an approximate (0,1,q) model using the techniques described by Box and Jenkins. Figure 5 shows the theoretical Allan variance curve for Cs 571 formed by converting the appropriate (1,1,1) model into an approximate (0,1,2) model. The actual values for the square root of the Allan variance for Cs 571 are also plotted. (The confidence limits are based upon the work of Audoin and Lesage.¹⁰) For longer sampling times the theoretical Allan variance overestimates the estimated Allan variance by a con-

siderable margin. This was true for most of the clocks in the study. The ability of simple ARIMA models to reproduce the observed Allan variance is thus limited.

Appendix 1

Let $y(t)$ represent the one-day average fractional frequency value at day t ; $\hat{y}_t(1)$ be the forecast of $y(t+1)$ made at day t ; and $a(t) = y(t) - \hat{y}_{t-1}(1)$ be the one-day forecast error at day t ; then, given fractional frequency values up to day t , the prediction equations for the (0,1,2) ARIMA model are:

$$\hat{y}_t(1) = y(t) + \theta_0 - \theta_1 a(t) - \theta_2 a(t-1)$$

$$\hat{y}_t(2) = \hat{y}_t(1) + \theta_0 - \theta_2 a(t)$$

$$\hat{y}_t(j) = \hat{y}_t(j-1) + \theta_0, \quad j = 3, 4, 5, \dots$$

If $\theta_2 = 0$ in the above equations, then the above equations become the prediction equations for the (0,1,1) ARIMA model. Likewise, the prediction equations for the (1,1,1) ARIMA model are:

$$\hat{y}_t(1) = (1 + \theta_1)y(t) - \theta_1 y(t-1) + \theta_0 - \theta_1 a(t)$$

$$\hat{y}_t(2) = (1 + \theta_1)\hat{y}_t(1) - \theta_1 y(t) + \theta_0$$

$$\hat{y}_t(j) = (1 + \theta_1)\hat{y}_t(j-1) - \theta_1 \hat{y}_t(j-2) + \theta_0, \quad j = 3, 4, 5, \dots$$

References

1. Allan, D.W., et al., "The National Bureau of Standards Atomic Time Scale: Generation, Stability, Accuracy, and Accessibility," in National Bureau of Standards Monograph 140, Time and Frequency: Theory and Fundamentals, edited by B.E. Blair, pp. 205-231 (May 1974).
2. Barnes, J.A., "Models for the Interpretation of Frequency Stability Measurements," National Bureau of Standards Technical Note 683 (August 1976).
3. Meditch, J.S., "Clock Error Models for Simulation and Estimation," Aerospace Report No. TOR-0076(64 74-01)-2, The Aerospace Corporation (July 1975).
4. Box, G.E.P., and Jenkins, G.M., Time Series Analysis: Forecasting and Control, Revised Edition, Holden-Day, San Francisco (1976).
5. Morrison, N., Introduction to Sequential Smoothing and Prediction, McGraw-Hill, New York (1969).
6. Rutman, J., and Ubersfeld, J., "A Model for Flicker Frequency Noise of Oscillators," Proc. IEEE, 60, pp. 233-235 (February 1972).
7. Percival, D.B., "A Heuristic Model for Long-Term Atomic Clock Behavior," Proceedings of the 30th Annual Symposium on Frequency Control, pp. 414-419 (1976).
8. Barnes, J.A., et al., "Characterization of Frequency Stability," IEEE Trans. Instrum. Measure, IM-20, pp. 105-120 (May 1971).
9. Conover, W.J., Practical Nonparametric Statistics, John Wiley & Sons Inc., New York (1971).
10. Lesage, P. and Audoin, C., "Characterization of Frequency Stability: Uncertainty due to the Finite Number of Measurements," IEEE Trans. Instrum. Measure, IM-22, pp. 157-161 (June 1973).

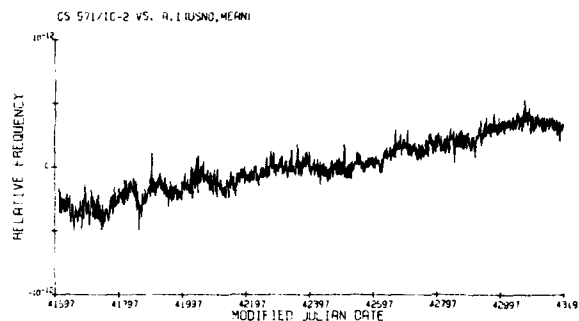


Figure 1: One Day Average Fractional Frequency Values For Cs 571 vs. A.1(USNO,MEAN)

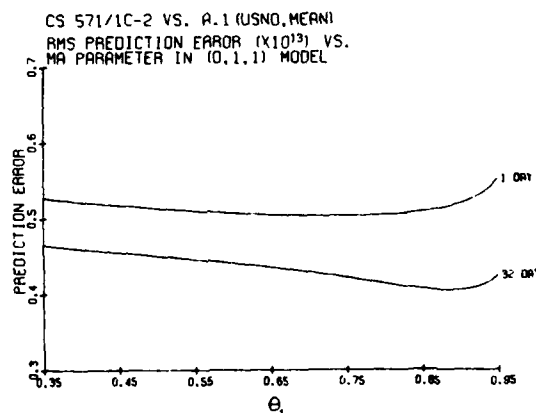


Figure 2:

RMS Prediction Error ($\times 10^{13}$) for Cs 571 vs. A.1(USNO,MEAN) vs. Moving Average Parameter in (0,1,1) Model

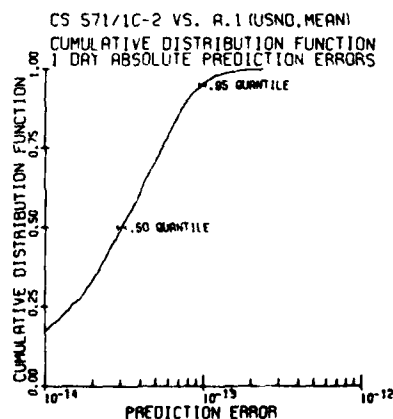


Figure 3: Cumulative Distribution Function of One Day Absolute Prediction Errors for Cs 571 vs. A.1(USNO,MEAN)

TABLE 1

OSCILLATOR	DAYS OF DATA	LINEAR FREQUENCY DRIFT	COMMENTS
Cs 346	1600	3.2×10^{-13} /year	
Cs 532	1600	not significant	
Cs 549	1600	not significant	
Cs 571	1587	1.6×10^{-13} /year	High performance beam tube
Cs 591	1338	3.9×10^{-13} /year	
Cs 654	1330	-5.1×10^{-13} /year	High performance beam tube
Cs 660	1352	not significant	High performance beam tube
Cs 783	1245	non-linear	High performance beam tube
Cs 834	1127	1.8×10^{-13} /year	High performance beam tube
Cs 837	1016	non-linear	High performance beam tube

TABLE 2

OSCILLATOR	MODEL	$\theta_0 (\times 10^{13})$	MODEL PARAMETERS	$\sqrt{\hat{\sigma}_a^2} (\times 10^{13})$	$\hat{\alpha}_p$	$\hat{\alpha}_w$
Cs 346	(0,1,1)	+0.011	$\hat{\theta}_1 = .86$	1.44	.13	.19
Cs 532	(0,1,1)	0	$\hat{\theta}_1 = .84$	1.20	.60	.10
Cs 549	(0,1,2)	0	$\hat{\theta}_1 = .75$ $\hat{\theta}_2 = .05$	1.24	.09	>.25
	(0,1,1)	0	$\hat{\theta}_1 = .80$	1.24	.02	.10
Cs 571	(1,1,1)	+0.003	$\hat{\theta}_1 = .90$ $\hat{\theta}_1 = .30$	0.49	.80	>.25
	(0,1,1)	+0.003	$\hat{\theta}_1 = .72$	0.50	<.01	<.001
Cs 591	(0,1,1)	+0.013	$\hat{\theta}_1 = .92$	1.61	.004	<.01
Cs 654	(0,1,2)	-.012	$\hat{\theta}_1 = .60$ $\hat{\theta}_2 = .15$	0.69	.15	>.25
	(0,1,1)	-.012	$\hat{\theta}_1 = .72$	0.70	<.01	<.001
Cs 660	(1,1,1)	0	$\hat{\theta}_1 = .80$ $\hat{\theta}_1 = .45$	0.81	.25	>.25
	(0,1,1)	0	$\hat{\theta}_1 = .41$	0.83	<.01	<.001
Cs 783	(0,1,1)	-.010	$\hat{\theta}_1 = .73$	0.56	.30	>.25
Cs 834	(1,1,1)	+0.007	$\hat{\theta}_1 = .90$ $\hat{\theta}_1 = .45$	0.54	>.995	>.25
	(0,1,1)	+0.007	$\hat{\theta}_1 = .54$	0.57	<.01	<.001
Cs 837	(0,1,1)	-.023	$\hat{\theta}_1 = .78$	0.93	<.01	<.001

$\hat{\alpha}_p$ is the critical level of the portmanteau lack of fit test; $\hat{\alpha}_w$ is the critical level of the cumulative periodogram test.

TABLE 3
RMS PREDICTION ERRORS FOR ARIMA MODELS ($\times 10^{13}$)

OSCILLATOR	ARIMA MODEL	LEAD TIMES (DAYS)							
		0	1	2	4	8	16	32	64
Cs 346	(0,1,1)	.011	1.44	1.11	0.92	0.75	0.71	0.73	0.90
	(0,1,1)	0	1.44	1.12	0.92	0.76	0.73	0.75	1.04
Cs 532	(0,1,1)	0	1.20	0.94	0.75	0.64	0.63	0.66	0.88
Cs 549	(0,1,2)	0	1.24	1.01	0.85	0.84	0.83	0.98	0.99
	(0,1,1)	0	1.24	1.01	0.85	0.84	0.83	0.99	0.99
Cs 571	(1,1,1)	.003	0.49	0.42	0.38	0.34	0.34	0.39	0.43
	(1,1,1)	0	0.50	0.42	0.38	0.34	0.34	0.40	0.47
	(0,1,1)	.003	0.50	0.43	0.39	0.35	0.37	0.42	0.50
	(0,1,1)	0	0.50	0.43	0.39	0.35	0.37	0.43	0.53
Cs 591	(0,1,1)	.013	1.61	1.23	0.95	0.78	0.72	0.60	0.45
	(0,1,1)	0	1.61	1.24	0.96	0.81	0.75	0.64	0.57
Cs 654	(0,1,2)	-.012	0.69	0.59	0.50	0.54	0.49	0.63	0.77
	(0,1,2)	0	0.69	0.60	0.50	0.55	0.53	0.67	0.89
	(0,1,1)	-.012	0.70	0.60	0.50	0.55	0.50	0.64	0.76
	(0,1,1)	0	0.70	0.60	0.51	0.56	0.53	0.68	0.87
Cs 660	(1,1,1)	0	0.81	0.78	0.75	0.74	0.83	1.02	1.41
	(0,1,1)	0	0.83	0.82	0.81	0.80	0.91	1.04	1.44
Cs 783	(0,1,1)	-.010	0.56	0.45	0.42	0.41	0.46	0.58	0.95
	(0,1,1)	0	0.56	0.46	0.42	0.41	0.47	0.63	1.05
Cs 834	(1,1,1)	.007	0.54	0.51	0.44	0.44	0.44	0.51	0.56
	(1,1,1)	0	0.54	0.51	0.45	0.44	0.46	0.54	0.58
	(0,1,1)	.007	0.57	0.54	0.49	0.53	0.52	0.56	0.65
	(0,1,1)	0	0.57	0.54	0.49	0.53	0.54	0.57	0.64
Cs 837	(0,1,1)	-.023	0.93	0.75	0.61	0.61	0.59	0.65	0.81
	(0,1,1)	0	0.93	0.75	0.62	0.63	0.62	0.79	0.98

TABLE 4
RMS PREDICTION ERRORS ($\times 10^{13}$) AND FILTER LENGTH (DAYS) FOR ORDER 1 FIXED FILTER

OSCILLATOR	LEAD TIME (DAYS)							
	1	2	4	8	16	32	64	
Cs 346	1.46	1.14	0.94	0.77	0.77	0.92	0.96	
	r=8	r=8	r=8	r=8	r=16	r=32	r=32	
Cs 532	1.20	0.94	0.75	0.68	0.67	0.77	0.83	
	r=8	r=8	r=8	r=16	r=16	r=32	r=32	
Cs 549	1.26	1.04	0.88	0.84	0.79	0.67	0.60	
	r=8	r=8	r=8	r=4	r=64	r=16	r=16	
Cs 571	0.53	0.45	0.40	0.37	0.36	0.38	0.49	
	r=4	r=8	r=8	r=8	r=8	r=8	r=64	
Cs 591	1.62	1.25	0.98	0.81	0.76	0.77	1.02	
	r=16	r=16	r=16	r=16	r=16	r=32	r=64	
Cs 654	0.72	0.61	0.54	0.45	0.60	0.61	0.78	
	r=8	r=8	r=4	r=4	r=16	r=16	r=16	
Cs 660	0.88	0.86	0.83	0.85	0.83	0.81	0.94	
	r=2	r=4	r=2	r=8	r=4	r=8	r=8	
Cs 783	0.58	0.47	0.43	0.43	0.54	0.63	1.10	
	r=4	r=4	r=8	r=8	r=2	r=2	r=2	
Cs 834	0.61	0.57	0.50	0.46	0.42	0.46	0.55	
	r=2	r=4	r=4	r=4	r=4	r=4	r=4	
Cs 837	0.96	0.78	0.65	0.60	0.69	0.74	1.13	
	r=4	r=8	r=8	r=4	r=4	r=16	r=16	

TABLE 5
RMS PREDICTION ERRORS ($\times 10^{13}$) AND FILTER LENGTH (DAYS) FOR ORDER 2 FIXED FILTER

OSCILLATOR	LEAD TIME (DAYS)							
	1	2	4	8	16	32	64	
Cs 346	1.49	1.19	1.02	0.90	0.91	1.04	1.43	
	r=32	r=32	r=64	r=64	r=64	r=64	r=128	
Cs 532	1.23	0.98	0.81	0.78	0.84	0.92	1.27	
	r=32	r=32	r=32	r=32	r=128	r=128	r=256	
Cs 549	1.30	1.07	0.92	0.84	0.77	0.74	0.83	
	r=32	r=256	r=256	r=256	r=256	r=256	r=256	
Cs 571	0.54	0.47	0.43	0.42	0.45	0.48	0.48	
	r=32	r=32	r=32	r=32	r=32	r=256	r=256	
Cs 591	1.61	1.27	0.94	0.76	0.71	0.68	0.67	
	r=128	r=128	r=128	r=128	r=128	r=128	r=128	
Cs 654	0.75	0.66	0.61	0.67	0.68	0.86	0.82	
	r=16	r=16	r=16	r=64	r=64	r=64	r=256	
Cs 660	0.99	0.94	0.91	0.95	1.11	1.33	1.82	
	r=16	r=32	r=32	r=32	r=64	r=64	r=128	
Cs 783	0.59	0.49	0.44	0.44	0.46	0.51	0.82	
	r=16	r=32	r=32	r=64	r=64	r=64	r=128	
Cs 834	0.64	0.59	0.54	0.51	0.54	0.65	0.83	
	r=32	r=32	r=32	r=64	r=64	r=64	r=128	
Cs 837	0.99	0.81	0.68	0.64	0.58	0.61	0.65	
	r=64	r=64	r=64	r=128	r=128	r=128	r=128	

TABLE 6

OSCILLATOR	$\sigma_y(\tau)$ ($\times 10^{13}$) AND RATIO OF BEST ARIMA MODEL TO $\sqrt{2}\sigma_y(\tau)$						
	τ (DAYS)						
	1	2	4	8	16	32	64
Cs 346	1.32	0.98	0.77	0.54	0.53	0.61	0.83
	1.30	1.25	1.18	1.02	1.06	1.18	1.30
Cs 532	1.09	0.83	0.59	0.49	0.48	0.54	0.73
	1.28	1.25	1.11	1.08	1.08	1.16	1.17
Cs 549	1.09	0.85	0.66	0.61	0.57	0.59	0.61
	1.24	1.19	1.10	1.03	0.97	0.85	0.87
Cs 571	0.41	0.35	0.30	0.26	0.25	0.30	0.28
	1.18	1.17	1.12	1.08	1.04	1.09	0.92
Cs 591	1.49	1.11	0.79	0.59	0.52	0.47	0.48
	1.24	1.28	1.18	1.07	1.02	1.11	1.51
Cs 654	0.58	0.49	0.38	0.39	0.39	0.45	0.56
	1.19	1.17	1.08	1.02	1.12	1.02	1.03
Cs 660	0.61	0.62	0.60	0.61	0.67	0.81	1.10
	1.07	1.12	1.13	1.17	1.14	1.12	1.10
Cs 783	0.49	0.36	0.31	0.30	0.40	0.63	1.10
	1.24	1.13	1.04	1.03	1.23	1.54	1.64
Cs 834	0.43	0.41	0.36	0.34	0.33	0.39	0.44
	1.13	1.14	1.16	1.09	1.06	1.08	1.11
Cs 837	0.79	0.62	0.47	0.44	0.42	0.51	0.82
	1.20	1.17	1.09	1.02	1.01	1.11	1.43

For each clock, the first row is $\sigma_y(\tau)$, and the second row is the ratio of the best ARIMA model to $\sqrt{2}\sigma_y(\tau)$.

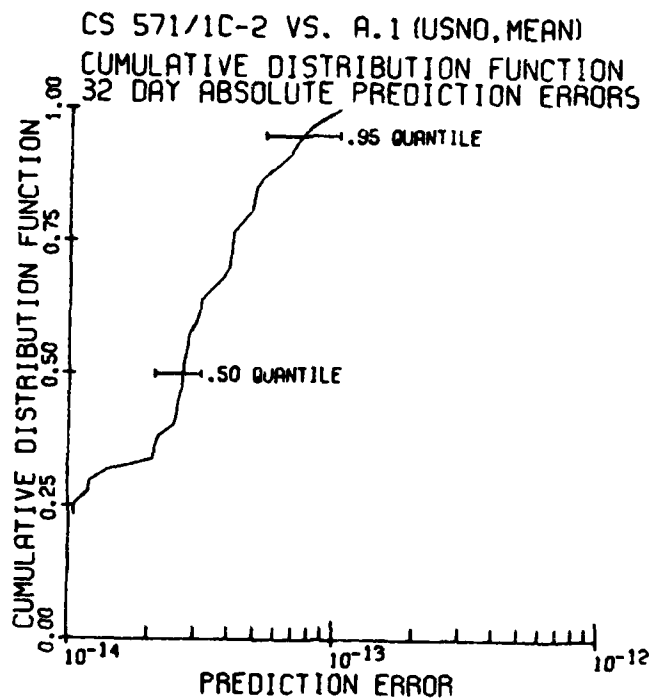


Figure 4: Cumulative Distribution Function of 32 Day Absolute Prediction Errors for Cs 571 vs. A.1(USNO,MEAN)

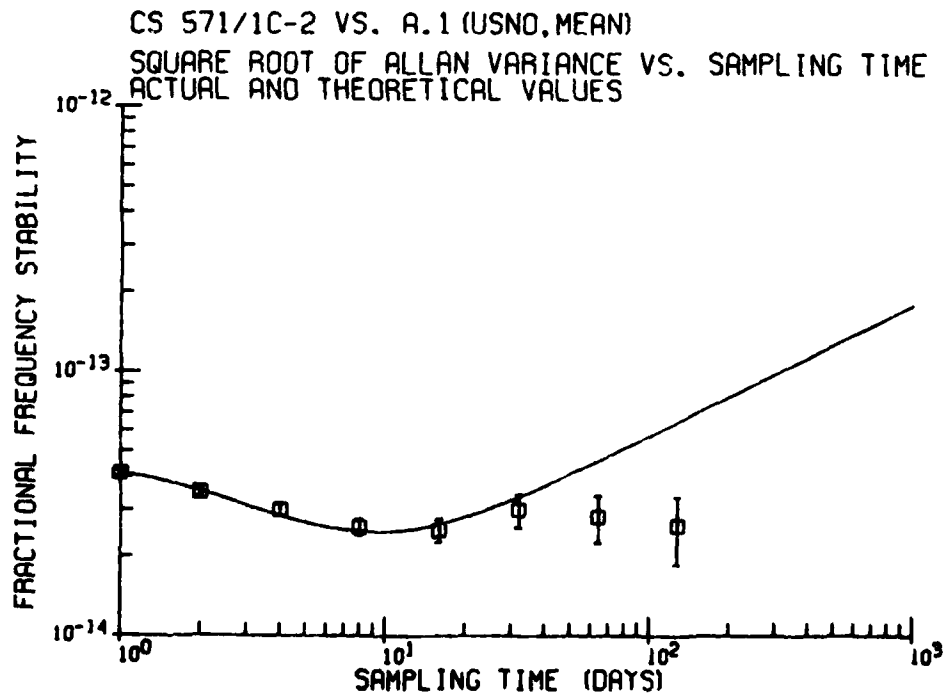


Figure 5: Actual and Theoretical Square Root of Allan Variance vs. Sampling Time for Cs 571 vs. A.1(USNO,MEAN)

MODELS AND PREDICTIONS FOR THE REALIZATION OF TIME SCALES

U. Hubner
Physikalisch-Technische Bundesanstalt Braunschweig
Federal Republic of Germany

Summary

As UTC(PTB) is a realization of the Co-ordinated Universal Time UTC, prediction of the time differences of these two scales is necessary. This has been discussed for lead times of 30 to 70 days. Up to now, no physical models of sufficient accuracy are available for the variations in the time differences of atomic clocks. Hence, stochastic models of the data are referred to for extrapolation. With regard to polynomials and specific clock models (autoregressive processes and ARIMA processes), the discussion shows that fitting a straight line is often preferred, rather than the more sophisticated models, particularly for the case when there are few data. One needs only 6 to 8 data of a time scale to extrapolate optimally with a fitted straight line.

Most time scales may be modelled very well (except one of the seven well known time scales discussed) by similar ARIMA (0,2,2)-processes which are nonstationary in second degree ($d = 2$). With the aid of these models, extrapolation takes place by a straight line adapting itself continuously to the given data. Thus, the excellent extrapolation by fitting a straight line is obvious and not trivial.

The analysis of the control mechanism for the realization of time scales shows that a further point of view should be considered: namely, the deviations of the extrapolated values from the real values ought to be less correlated and should have a symmetrical probability density around zero. If this is not the case, the realization will deviate on an average from the time scale to be realized.

1. Introduction

The Physikalisch-Technische Bundesanstalt (PTB) generates its time scale UTC(PTB) from its atomic time scale TA(PTB) by adding suitable corrections. These corrections are chosen such that UTC(PTB) differs as little as possible from UTC without following the short-lasting frequency fluctuations of UTC. In order to calculate the corrections, one needs the time differences TA(PTB)-TAI of the respective present time. PTB receives these data from the "Bureau International de l'Heure" (BIH) with a delay of 30 to 70 days. Consequently, extrapolations with these lead times are necessary. Up to now these extrapolations were carried out by fitting a straight line to the last $n = 5$ time differences known. This paper tries to establish particularly suited data models for the time differences, and to calculate optimum predicted values from the models.

It seems little expedient to apply the physical models which are successful in the characterization of atomic clocks, such as the white noise, the flicker noise, etc., to the time differences of time scales as these types of noise represent in each case only individual terms out of the serial development of the spectral density of the stochastic process. Instead, non-stationary models from the general class of the "autoregressive integrated moving average" ARIMA processes are treated including a short discussion of the special case of the stationary autoregressive processes as a model.

2. ARIMA models

ARIMA processes z_t are generated by filtering a white noise process a_t . The filtering is defined by

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d z_t = \theta_0 + (1 - \theta_1 B - \dots - \theta_q B^q) a_t.$$

t is a discrete variable, as e.g. the time, and B is the backward shift operator with $B.z_t = z_{t-1}$ or $B.a_t = a_{t-1}$; p or q are the orders of the autoregressive or moving average operators. The parameters ϕ_i ($i = 1, 2, \dots, p$) and θ_k ($k = 1, 2, \dots, q$) as well as

the variance σ_a^2 of the white noise a_t represent degrees of freedom for the modelling of a specified stochastic process. On the other hand, the case $\theta_0 \neq 0$ in connection with the choice of a value for d produces a certain "deterministic" trend in the form of a polynomial of degree d . For many applications, the assumption of a "stochastic trend" is more realistic than the assumption of a deterministic trend, since the stochastic trend follows a changing of the general trend of the given data. Stochastic trends are defined by the factor $(1-B)^d$ which means a polynomial of degree $d-1$ adjusting continuously to the given data.

The pattern of the time differences TAI-TA(PTB) which are to be modelled is shown in Figure 1 for the period of the past 8 years. The time differences are given discretely for periods of 10 days so that 292 data are available. Now one has to establish one or more ARIMA(p, d, q)-models for this stochastic process by suitably choosing the parameters p, d and q and by adjusting the parameters $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \theta_0$ and σ_a^2 .

A rough reduction of multiplicity of the possible models is provided by the following consideration: An unambiguous deterministic trend is not recognizable.

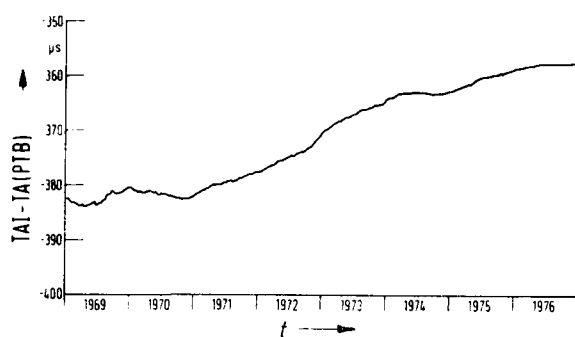


Fig. 1 TAI-TA(PTB) for the period from 1969 until 1976. The time differences are discretely submitted 10 days spaced.

Hence, such models will be particularly suited where Θ_0 is dispensed with, that is to say, the models should be non-stationary ($d > 0$). The absence of a clear deterministic trend is valid not only for the mean value but for the slope, too. To reproduce stochastic changes of the slope, $d \geq 2$ has to be chosen. Then the 2nd or higher differences of succeeding data form a stationary process.

Figure 2 includes the 1st and the 2nd differences of the given time differences. The mean of small data parts of the 1st differences is still subject to random changes whereas the 2nd differences seem to form a stationary process. Therefore $d = 2$ will be sufficient.

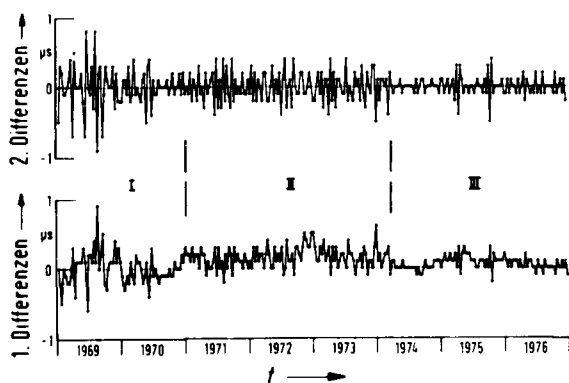


Fig. 2 1st and 2nd differences of the time differences plotted in Fig. 1. Differences are taken from succeeding data. The data parts I, II, III are to be represented by different stochastic models.

Moreover one recognizes from the 1st and 2nd differences that the time differences obviously belong to three different stochastic processes (parts I, II, III). The data of the years 1969 and 1970 show very high fluctuations which can be explained by the small number of clocks

available at the PTB at that time. Parts II and III differ distinctly from I. The data of part III are much more correlated than those of II. The mean of the 1st differences is smaller by nearly $0.1 \mu s$. This corresponds to a fractional frequency change of TA(PTB) against TAI by about $1 \cdot 10^{-13}$. The difference is substantial as is shown by the precise model adjustment described below.

No explanation could be found for this abrupt change of the stochastic qualities. At the time of this change neither modifications in generating TA(PTB) were made in the PTB nor were changes observed in the behaviour of the clocks. The reason can neither be found in changes of the calculation method of TAI (e.g. introduction of the ALGOS method) because the time differences between other important time scales and TAI do not show such a behaviour.

An important part in the reduction of the number of the possible models is played by the auto-correlation function (acf) and the partial auto-correlation function (pacf) of the given stochastic process. The pacf is established by fitting autoregressive processes with increasing order k to the given process. In each case, the last coefficient of the fitted autoregressive process equals the k -th value of the pacf. If the process is autoregressive of the order p , then the pacf equals zero for $k > p$. Thus the pacf serves to estimate the autoregressive character of the process. On the other hand, information can be obtained from the acf concerning the moving average character as well as a possible nonstationarity of the process.

Fig. 3 shows acf and pacf for the given data of part III as well as for their 1st and 2nd differences. The plotted limits L mean standard deviations of the acf and pacf. For the acf, the limits were calculated according to Bartlett's formula

$$s(r_k) \approx [1 + 2(r_1^2 + r_2^2 + \dots + r_q^2)]^{1/2} / N^{1/2}$$

$s(r_k)$ being the standard deviation of r_k 's of the acf for lags $k > q$; N being the number of the data from which r_k was calculated; q was set to 2 as the first two values of the acf show an irregular behaviour. The standard deviation of the pacf was estimated according to Quérouille's approximation $s(\hat{\phi}_{kk}) \approx 1/N^{1/2}$ where $\hat{\phi}_{kk}$ is the k -th value of the pacf. This approximation is valid for $k > p$.

The course of acf and pacf of the 1st differences indicates ARIMA(1,1,2)- or ARIMA(2,1,2)-models. However, a further forming of differences would be better because the acf descends slowly. The acf and the pacf of the 2nd differences then show the typical behaviour of stationary processes. According to this, the models (0,2,2), (0,2,3) and (1,2,1) seem to be suitable.

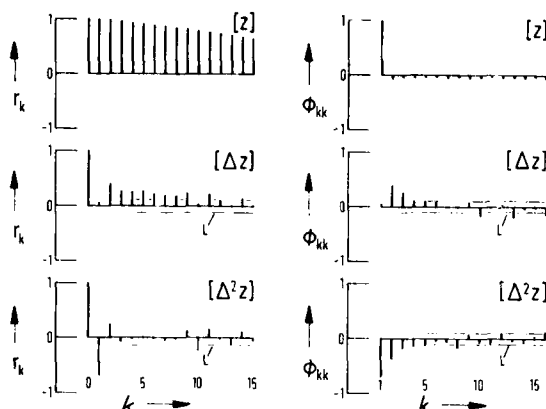


Fig. 3 Autocorrelation (r_k) and partial autocorrelation (ϕ_{kk}) functions of the time differences TAI-TA (PTB) of data part III, and their 1st and 2nd differences for lags up to $k = 15$.

The corresponding Figure of the data part II differs from those of part III particularly in the 1st differences. The acf as well as the pacf are distinctly smaller than the limits L , so that the simple model (0,1,0) should be very suitable.

The final adjustment of the parameters is on principle done by maximization of the log likelihood function.

$$l(\underline{\phi}, \underline{\theta}, \underline{\sigma}_a) = f(\underline{\phi}, \underline{\theta}) - N \cdot \ln \sigma_a - \frac{1}{2\sigma_a^2} S(\underline{\phi}, \underline{\theta})$$

$l(\underline{\phi}, \underline{\theta}, \underline{\sigma}_a)$ may be interpreted as a measure of the probability with which a set of $\underline{\phi}$ and $\underline{\theta}$ matches the set z of the given data. The shortened form $\underline{\phi}$ comprises all ϕ_i ($i = 1, \dots, p$), and $\underline{\theta}$ comprises all θ_k ($k = 1, \dots, q$). $S(\underline{\phi}, \underline{\theta})$ is the sum

$$S(\underline{\phi}, \underline{\theta}) = \sum_{t=-\infty}^N [a_t | \underline{\phi}, \underline{\theta}, z]^2$$

where $[a_t | \underline{\phi}, \underline{\theta}, z]$ denotes the expectation of a_t conditional on $\underline{\phi}$, $\underline{\theta}$ and z . The function $f(\underline{\phi}, \underline{\theta})$ influences l only in the case of small N . For moderate and large values of N , $l(\underline{\phi}, \underline{\theta}, \underline{\sigma}_a)$ is dominated by the sum S . Therefore minimization of S usually provides a very good maximization of the likelihood function. This is called the least squares estimation of parameters and was applied throughout in this paper. In lieu of the z_t , d -th differences w_t forming a stationary process are inserted if the z_t themselves are not stationary. The summation over t certainly extends to $-\infty$, but in practice the terms converge, for $t \ll 1$, rather rapidly against zero, if the parameters ϕ_i do not approach the value 1 which characterizes nonstationary behaviour. The estimation of the conditional expectations $[a_t | \underline{\phi}, \underline{\theta}, z]$ may then be made iteratively. After minimization of $S(\underline{\phi}, \underline{\theta})$, the residuals a_t approach well the stochastic characterization "white" and have minimized variance σ_a^2 if the dimensions p, d , and q of the model are suitably chosen. The minimization of S was performed with the algorithm of Marquardt, described in 2, which is based on a linearization of the dependence of the a_t from $\underline{\phi}$ and $\underline{\theta}$.

The ARIMA (p, d, q)-models which are possible according to the preliminary discussion were fitted by the least squares method. The results have been summarized in table 1 inclusive of the quantities σ_a^2 , Θ_0 , $P(\chi^2, \nu)$, and the uncertainties of the parameters important for a critical investigation. χ^2 and ν belong to the well known χ^2 -distribution, ν being the number for the degrees of freedom of the model adjustment. χ^2 is obtained from the autocorrelations r_a of the residuals a_t according to

$$\chi^2 = N \sum_{k=1}^K r_a^2(k),$$

where K is a maximum lag suitably chosen. ν then follows from $\nu = K - p - q - M$. M equals 1 or 0, depending on whether or not Θ_0 was taken into account as a parameter during the minimization of S . On the assumption that the χ^2 -distribution is valid, the quantity $P(\chi^2, \nu)$ indicates the probability (in percent) of the hypothesis that the value of χ^2 could have turned out larger than the calculated one. Thus the quantity P may be interpreted as a measure of the probability with which the just chosen model is in the vicinity of the ideally suitable model. The uncertainties of the parameters resulted from the roots of the diagonal elements of the parameter covariance matrix which is customary if one employs the least squares method. They must be considered as standard errors.

Table 1 shows that the parts II and III are stochastically different. While part II can be described with high probability (P) by a nonstationary model of 1st order, it proves for part III that $d = 1$ will be completely unsuitable. Obviously part II can be described very well by models of 1st and 2nd order, whereas models of 2nd or 3rd order are suitable for part III. Consequently the complete data set II + III cannot be represented by one very well fitting model. The differences become also evident in the variances σ_a^2 , whose values for parts II and III have nearly the ratio 2:1.

The different models of table 1 are partly equivalent, as e.g. autoregressive operators of the type $(1-\Theta B)$ are represented as an infinite series on the moving average side of the recursive filter equation and may be interpreted as moving average parameters. The same applies vice versa for moving average operators of the type $(1-\Theta B)$.

In table 2, optimum estimated ARIMA-models of some well known time scales have been summarized. These models refer to the data part of the homogenous stochastic character which contains in each case the latest data (up to the end of 1976 inclusive). All data of the time scale TA(USNO) could be used because the stochastic character alone of this time scale has remained unchanged during 9 years. All other time scales reduced their initially strong starting noise in the course of time to values of the

same order as TAI-TA(USNO). Therefore column 2 of the table comprises the amount of data used in each case. Only 36 data of the time scale TA(NRC) could be used as the 1st differences of TAI-TA(NRC) show a fractional frequency step of $1 \cdot 10^{-12}$ at the turn of the year 1975 to 1976.

The general stochastic structure of the time differences TAI-TA(i) is obviously very well approximated by a (0,2,2)-model. This is approximately valid also for TAI-TA(PTB). The (0,2,2)-model of these time differences has yet a large value of probability $P(x^2, v)$ which is equal to 42%. The only exception are the time differences TAI-TA(ON) probably containing periodic fluctuations. For periodic processes, one has to employ the so-called seasonal ARIMA models; however, it is not intended to discuss them in this paper.

An interesting conclusion can be drawn from the fact that the time differences TAI-TA(i) are generally well described by (0,2,2)-models. The spectral density $p(f)$ of the second differences considered as stationary in this model can be calculated directly from their (0,0,2)-model.

$$p(f) = 2\sigma_a^2 [1 + \theta_1^2 + \theta_2^2 - 2\theta_1(1 - \theta_2)\cos(2\pi f) - 2\theta_2\cos(4\pi f)].$$

f means a normalized frequency for which $0 \leq f \leq 1/2$ is valid. The maximum frequency $f = 1/2$ corresponds to the $(N/2)$ -fold of the fundamental frequency $(N \cdot 10d)^{-1}$, N being the number of the given data, and 10 days (d) the data spacing. The second differences may be considered approximately as differentials of the first differences. Thus one obtains the spectral density $p(f)/(4\pi^2 f^2)$ for the first differences. Writing the Taylor series for the cosinus function ($\cos x = 1 - x^2/2! + x^4/4! - \dots$) proves that the series of the spectral density for the first differences contains only the even powers f^{-2}, f^0, f^2, \dots of the frequency. These powers correspond to different noise types in the $\sigma(\tau)$ -diagram which can be calculated from the spectral density (e.g. Barnes⁵). $\sigma(\tau)$ means the so-called two sample variance. It contains the terms of "random walk" (f^{-2}), "white frequency noise" (f^0), "white phase noise" (f^2), and so on, however, no odd powers like the term in f^{-1} corresponding to the flicker noise. Obviously the time differences TAI-TA(i) contain few or no parts at all of the noise type called flicker noise of the frequency.

3. Extrapolation

In the case of ARIMA processes, the optimum prediction is obtained by calculating the expectation of the requested quantity. At time t (origin of prediction), $z_t(1)$ is set as the optimum forecast of the stochastic variable z_t for the time

$t + l$ (l = lead time). Then $z_t(1) = E[z_{t+l} | z_t, z_{t-1}, \dots]$, which means that $z_t(1)$ equals the expectation of z_{t+l} conditional on z_t, z_{t-1}, \dots which have already occurred. In the following, this rule is applied to three equivalent forms of ARIMA-models (see remarks at the end of chapter 2 concerning the equivalence of models) in order to get information on extrapolation errors, uncertainties and correlations.

(a) $z_t(1)$ can be interpreted as an infinite weighted sum of previous observations:

$$z_t(l) = \pi_1^{(l)} z_t + \pi_2^{(l)} z_{t-1} + \dots$$

The weights $\pi_1^{(l)}$ and $\pi_2^{(l)}$ of the ARIMA (1,2,1)-model of part III were plotted in Figure 4 as examples. The contribution of the weights of previous observations decreases rapidly. Moreover, the larger the lead time l is, the stronger the last known value of z_t contributes to $z_t(1)$.

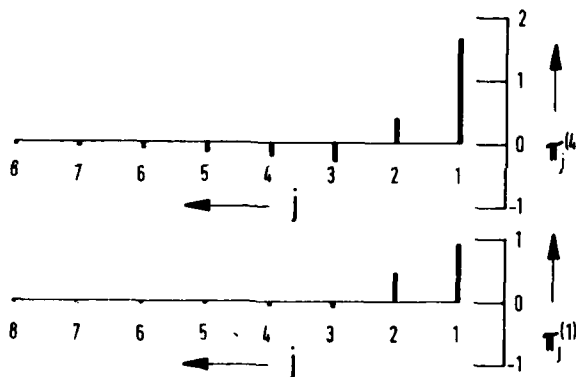


Fig. 4 The weights of previous data for the extrapolation with origin t over lead times of 1 ($\pi_1^{(1)}$) or 4 ($\pi_1^{(4)}$) in units of 10 days.

(b) By remodelling $z_t(1)$ to the weighted sum

$$z_{t+l} = a_{t+l} + \psi_1 a_{t+l-1} + \psi_2 a_{t+l-2} + \dots,$$

one obtains from the expectation of the square of z_{t+l} , a statement on the uncertainty of a prediction. For the (0,2,2)-model being very well suited for the time differences TAI-TA(i), the so-called variance function may be calculated

$$V(l) = 6\sigma_a^2 [1 + (l-1)\lambda_0^2 + \frac{1}{6}l(l-1)(2l-1)\lambda_1^2 + \lambda_0\lambda_1l(l-1)]$$

with $\lambda_0 = 1 + \theta_2$ and $\lambda_1 = 1 - \theta_1 - \theta_2$. Then the root of $V(1)$ is the standard deviation $s_E(1)$ of the extrapolation error for extrapolation 1 steps ahead. In Figure 5, $s_E(1)$ is plotted as a function of l for the (0,2,2)-models of the institutes indicated in table 2. The extrapolation over lead times of 30 to 70 days which is required for the calculation of UTC(PTB) therefore contains uncertainties of 0.15 to 0.3 μ s in case of optimum extrapolation.

This is about the order of magnitude of the uncertainties caused by the time transmission method LORAN-C in the time differences. Improvement of the mathematical extrapolation method thus yields no further reduction of the extrapolation uncertainties.

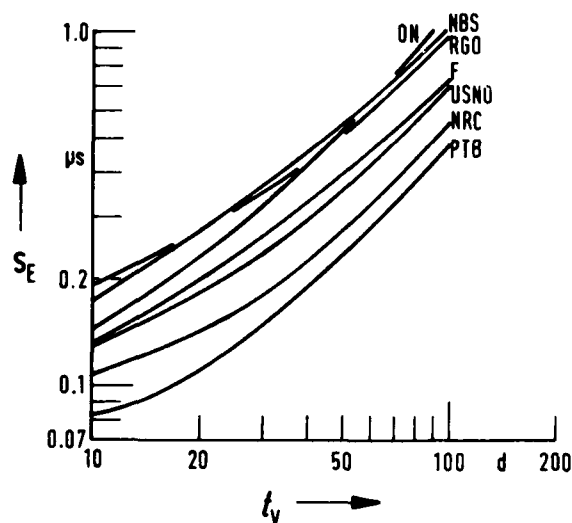


Fig. 5 Standard deviations s_E of the extrapolation error as a function of the lead time t_v for the time differences TAI-TA(i) of some well known time scales (i).

(c) The extrapolated value $z_t(1)$ is obtained in the simplest way from the recursive definition of the ARIMA-processes (cf. beginning of chapter 2). Accordingly, the extrapolated values of the (0,2,2)-model result from the equations

$$z_t(1) = 2z_t - z_{t-1} - \Theta_1 a_t - \Theta_2 a_{t-1},$$

$$z_t(2) = 2z_t(1) - z_t - \Theta_2 a_t,$$

$$z_t(l) = 2z_t(l-1) - z_t(l-2); \quad l > 2.$$

For the times t , the residuals a_t continuously result from $a_t = z_t - z_{t-1}(1)$

The first extrapolations may be executed with the initial values $a_t = 0$. The error caused by this approximation reduces rather rapidly with increasing time t . However, in the following examples a_t 's were employed which resulted from the model adjustment according to the least squares method. Therefore the examples do not include any initial errors.

In the cited recursive form of the extrapolation process, the parameter Θ_0 of the deterministic trend was set to zero. On account of the stochastic structure of the time differences, it is necessary to dispense with deterministic trends. Let us suppose that only the data set of part II of the PTB time differences is known whose first

differences have the mean value

$\Delta z(\text{II}) = 0,156 \mu\text{s}$. Then a (0,1,0)-model which is very well suited for this part has $\Theta_0 = \Delta z = 0,156 \mu\text{s}$. This value corresponds in part II to a mean deterministic trend of the fractional frequency differences of $0,156 \mu\text{s}/10 \text{ d} \approx 1.8 \times 10^{-13}$. If the times $t+1$ for which extrapolated values are calculated, lie within II, then the deterministic (0,1,0)-model employing $\Theta_0 = 0,156 \mu\text{s}$ yields very good extrapolated values. However, in 1974, there was a change of the stochastic character of the time differences. The mean of the 1st differences now takes the value $0,056 \mu\text{s}$ for part III. The deterministic (0,1,0)-model does not follow this change of the mean fractional frequency difference of 1.1×10^{-13} , but it interprets this change for ever as an extrapolation error. As there is no important physical reason to assume a deterministic trend (with e.g. $\Theta_0 = 0,156 \mu\text{s}$), it would be better to allow only stochastic trends for the frequency differences. This is achieved by selecting models with $d \geq 2$. For $d = 2$ this is seen from the third of the three recursive extrapolation equations which is valid for $l > 2$. It represents the recursive form of a straight line which passes through $z_t(1)$ and $z_t(2)$. However, $z_t(1)$ and $z_t(2)$ have as contributory determinants the observed residuals a_t, a_{t-1} , and so on. Thus this form of extrapolation adapts the slope of its extrapolation straight line (or the mean of its frequency differences) to the new conditions of part III.

A further criterion for the possibility to use a chosen model for extrapolation is based on the autocorrelation function of the extrapolation errors. Models which yield highly correlated extrapolation errors are unqualified for extrapolation in the UTC(PTB)-calculation. This UTC(PTB)-calculation contains the extrapolation error as a correction to the preceding UTC-UTC(PTB) difference. Thus, extrapolation errors having for a long period the same sign sum up to an increasing value. This leads to large deviations of the time differences from the wanted value zero. Figure 6 comprises the autocorrelation functions of the extrapolation errors calculated for the two models (0,1,0) and (2,1,2) of the data part II using the deterministic trends ($\Theta_0 \neq 0$). According to this, the (2,1,2)-model combined with the control of the UTC(PTB)-calculation¹ is better suited than the (0,1,0)-model though the latter model realizes the stochastic process very well ($P(\chi^2, \nu) = 64 \%$).

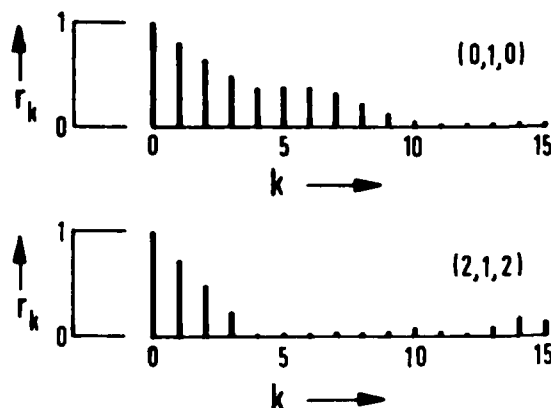


Fig. 6 Autocorrelation functions of extrapolation errors of the data part II for the time differences TAI-TA(PTB) represented by the ARIMA-models (0,1,0) and (2,1,2). Though the model (0,1,0) represents the given time differences very well ($P(\chi^2, \nu) = 64\%$), the model (2,1,2) is better suited for the calculation of UTC(PTB) because of smaller correlations of the extrapolation errors.

An interesting conclusion can be drawn from the fact that time scale differences are usually well represented by (0,2,2)-models for an optimum extrapolation from few data available. Figure 7 compares the empirically estimated standard deviations s_E of the extrapolation errors of different extrapolation methods for the data parts II, III, and II + III of the time differences TAI-TA(PTB). The extrapolation was performed for a lead time of 40 days (data spacing $l = 40$ d); $l = 40$ d is typical of the extrapolation problem in the UTC(PTB)-calculation. In Figure 7, p means the number of time differences used to fit the straight line or the parabola. In the case of the merely autoregressive process, p is its order.

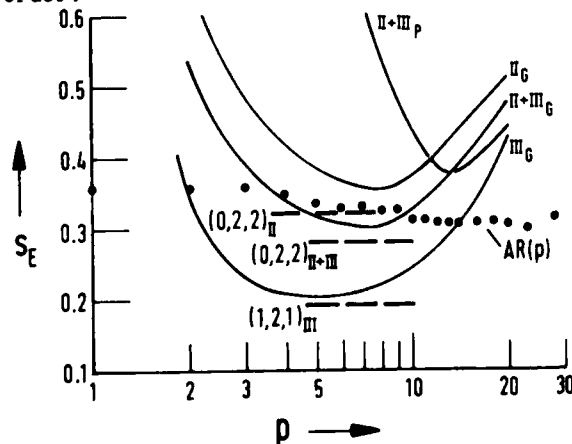


Fig. 7 The standard deviations of the extrapolation error for different extrapolation methods: method of fitting a straight line (index "G"), of fitting a parabola (index "p"), of fitting autoregressive processes (AR(p)) of the order p , and method of ARIMA-processes. Here p is the number of the data used for fitting the straight line and the parabola. In case of the method of the autoregressive processes, however, p is its order. For comparison reasons, s_E was calculated and plotted for the data parts II, III, and II + III. Merely the autoregressive processes were fitted only to the whole data set of II + III, and the pertinent extrapolation error was calculated.

It turns out that the extrapolation performed by fitting a straight line for $p \approx 6$ to 8 is practically equivalent to the optimized ARIMA-models. The reason is not trivial and is based on the fact that the nonstationary process of the given data is particularly well represented by an ARIMA-model with $d = 2$. This ARIMA-model has - as mentioned above - a dominating stochastic trend in the form of a continuously adapting polynomial of the degree $d - 1 = 1$, that is to say, a straight line. Thus the fitted straight line extrapolates in a similar manner as the optimum adapted ARIMA-process. This is also confirmed by the fact that the autocorrelation functions of the extrapolation error practically do not differ for the methods of a fitted straight line and the (0,2,2)-model. Consequently, an optimum extrapolation can in general be performed by the aid of the pertinent fitted straight line if 6 to 8 time differences TAI-TA(i) are known.

For a further comparison of the extrapolation methods, two additional methods were employed, namely the extrapolation with the aid of a fitted parabola and the extrapolation with an autoregressive process fitted to the given data. The comparison shows that both methods are obviously less appropriate. The method of fitting a parabola yields too large standard deviations of the extrapolation error. The autoregressive process is too large-scaled for application. Moreover, it is in the main only a special case of the ARIMA-processes if one sets $d = q = 0$. Doing this, many parameters become automatically necessary for adaption purposes. In this paper, the necessary number p of parameters was estimated from the behaviour of the variance 6% of the residuals. This variance obviously tends towards an inferior limit so that $p \approx 70$ was estimated. The estimation of such a large number of parameters was not possible because of the small amount of data available. Therefore the estimation had to be made according to the behaviour of the autoregressive models up to the order $p \approx 25$.

References

- 1 U. Hübner: Der Algorithmus zur Berechnung der Amtlichen Zeitskala; PTB-Mitteilungen (1977); submitted for publication
- 2 G.E.P. Box and G.M. Jenkins: Time Series Analysis Forecasting and Control; (Holden-Day, San Francisco, USA (1970))
- 3 J.A. Barnes et al: Characterization of Frequency Stability; IEEE IM-20 (1971), 105

part	p	d	q	Φ	Θ	θ_0 in μs	σ_a in μs	$P(\chi^2, \nu)$ in %
II	0	1	0	-	-	0,156	0,153	64
	2	1	2	$\Phi_1 = 1,75 \pm 0,07$ $\Phi_2 = -0,84 \pm 0,07$	$\theta_1 = 1,80 \pm 0,04$ $\theta_2 = -0,94 \pm 0,04$	0,0142	0,150	88
	0	2	2	-	$\theta_1 = 1,00 \pm 0,10$ $\theta_2 = -0,08 \pm 0,10$	0,000524	0,154	86
	1	2	1	$\Phi_1 = -0,09 \pm 0,1$	$\theta_1 = 0,91 \pm 0,04$	0,00057	0,154	87
III	0	1	0	-	-	0,056	0,093	<0,1
	2	1	2	$\Phi_1 = 0,57 \pm 0,27$ $\Phi_2 = 0,33 \pm 0,27$	$\theta_1 = 0,71 \pm 0,29$ $\theta_2 = -0,07 \pm 0,26$	0,0045	0,082	45
	0	2	2	-	$\theta_1 = 1,07 \pm 0,10$ $\theta_2 = -0,30 \pm 0,10$	0,000084	0,083	42
	1	2	1	$\Phi_1 = -0,41 \pm 0,11$	$\theta_1 = 0,70 \pm 0,08$	0,000120	0,083	55
	2	3	2	$\Phi_1 = -0,40 \pm 0,002$ $\Phi_2 = 0,02 \pm 0,10$	$\theta_1 = 1,70 \pm 0,08$ $\theta_2 = -0,70 \pm 0,08$	-0,000088	0,085	50
II+III	0	2	2	-	$\theta_1 = 1,00 \pm 0,07$ $\theta_2 = -0,15 \pm 0,07$	-0,000469	0,126	33
	1	2	1	$\Phi_1 = -0,17 \pm 0,08$	$\theta_1 = 0,82 \pm 0,04$	-0,000477	0,126	36

Table 1 Comparison of the parameters of ARIMA-models for the data parts II, III and II + III of the time differences TAI-TA(PTB). p, d and q are the orders of the autoregressive operators, of the order of differenciation and of the order of the moving average operator; Φ is the set of the used autoregressive parameters; Θ is the set of the used moving average parameters. θ_0 characterizes a deterministic trend contained in the given data. σ_a is the standard deviation of the residual noise a_t . $P(\chi^2, \nu)$ was taken from the χ^2 -distribution, ν being the number of the degrees of freedom of the used model. $P(\chi^2, \nu)$ may be taken as the probability of how well the model represents the given process.

Institute	N	p	d	q	ϕ	or	θ	θ_0 in μs	σ_a in μs	$P(\chi^2, \nu)$ in %
F	200	0	2	2	$\theta_1 = 0,856 \pm 0,07$		$\theta_2 = -0,016 \pm 0,07$	-0,00141	0,132	78
NBS	120	0	2	2	$\theta_1 = 0,785 \pm 0,09$		$\theta_2 = 0,047 \pm 0,09$	-0,00665	0,174	67
NRC	36	0	2	2	$\theta_1 = 1,150 \pm 0,16$		$\theta_2 = -0,341 \pm 0,16$	0,00122	0,108	80
ON	150	0	2	2	$\theta_1 = 0,706 \pm 0,08$		$\theta_2 = -0,009 \pm 0,08$	0,00113	0,142	24
PTB(III)	100	0	2	2	$\theta_1 = 1,069 \pm 0,10$		$\theta_2 = -0,301 \pm 0,10$	0,00008	0,083	42
	100	1	2	1	$\phi_1 = -0,407 \pm 0,11$		$\theta_1 = 0,700 \pm 0,09$	0,00012	0,083	55
RG0	177	0	2	2	$\theta_1 = 1,000 \pm 0,08$		$\theta_2 = -0,144 \pm 0,08$	-0,00048	0,193	69
USNO	292	0	2	2	$\theta_1 = 1,020 \pm 0,06$		$\theta_2 = -0,200 \pm 0,06$	-0,00062	0,130	76

Table 2 Comparison of the parameters of the best suited ARIMA-models for the time differences TAI-TA(i) of some well known time scales. N is the number of the used data. In each case, the latest data set of homogenous character was taken. Further explanations as in table 1.

ACCURATE MEASUREMENTS OF SPECTRAL DENSITY OF PHASE NOISE IN DEVICES

F. L. Walls and S. R. Stein

Frequency and Time Standards Section
National Bureau of Standards
Boulder, CO 80302
(303) 499-1000, extension 3207

ABSTRACT

Systematic errors larger than 10 dB can occur in the measurement of the spectral density of phase unless considerable caution is exercised. Some potential problems due to the shape of the analyzer passband and the Fourier frequency dependence of mixers are discussed. Three measurement systems are analyzed to determine the conditions under which they may be used to make spectral density measurements with an accuracy of 0.2 dB.

INTRODUCTION

The systematic errors in the measurement of the spectral density of phase which we discuss concern the shape of the analyzer passband and the Fourier frequency dependence of double-balanced mixers. Both of these problems are well known. However, careful calculations and measurements indicate that in certain regions of the Fourier frequency spectrum the potential errors are substantially larger than had been expected. With respect to the wave analyzer passband, we find that errors larger than 10 dB can occur when making measurements at a frequency of three times the noise bandwidth of the analyzer when the spectral density diverges as rapidly as f^{-5} . We therefore present three figures which show the potential error due to the width and shape of the analyzer filter as a function of Fourier frequency. Curves are given for 2, 4, 6, 8 and 16 pole filters and for spectral densities which diverge as f^{-1} , f^{-3} , and f^{-5} so that the region of significant error may be defined for a wide variety of measurement situations.

Measurements have been made to characterize the performance of double-balanced mixers in detail because it has been found that the phase to voltage conversion sensitivity of these devices can vary 15 dB over a Fourier frequency range smaller than 0.1% of the nominal bandwidth. Several figures are included which demonstrate the dependence of such mixers on the input drive levels, the input impedances, and the output termination. Based on this data, we analyze specific techniques for obtaining high accuracy measurements. This discussion stresses wide

applicability, calibration accuracy and convenience, noise floor, and immunity from spurious effects. One technique stands out from the remainder because it is equal to or superior to the others in all respects. By using the methods described, we conclude that convenient measurements of spectral density of phase may be made to an accuracy of 0.2 dB.

FILTER SHAPES

In order to actually measure the spectral density of phase one would need a delta function filter. In this section, we present three graphs which quantify the errors which one would make in approximating the delta function by a real filter having a bandpass shape which is equivalent to $n/2$ high pass poles and an equal number of low pass poles. The specific function which has been assumed for the purpose of these calculations is:

$$\text{relative power} = \frac{1}{[1 + (\Delta f)^2]^{n/2}}$$

where Δf is the frequency difference from the center of the filter. When defined in this simple manner, the noise bandwidth of the filter is a function of the number of poles. The noise bandwidths for $n = 2, 4, 6, 8$ and 16 are listed in table I.

TABLE I

Number of poles, n	Noise bandwidth, Hz
2	3.14
4	1.57
6	1.18
8	0.983
16	0.631

The most common technique for measuring the spectral density of phase involves applying a signal voltage and a reference voltage to a double-balanced mixer and analyzing the mixer output. The noise voltage measured at the IF port is simply related to the phase fluctuations between the two inputs provided they are very near phase quadrature and the total phase fluctuations are small compared to 1 radian. If two oscillators are

being compared, a loose phase-lock loop is often used to maintain the quadrature condition. However, the transfer function of such a loop will have the effect of high pass filtering the phase fluctuations [1]. Even if a phase-lock loop is not employed, it is often necessary to high pass filter the noise in order to obtain sufficient amplification to measure low noise levels. Consequently we have assumed, for the purpose of this calculation, that the spectral density being measured varies as:

$$S_{\phi}(f) = [1 + (f/0.25 \text{ Hz})^2]^{-p/2},$$

for $p = 1, 3$ and 5 .

Figures 1, 2 and 3 show the relative error in estimating such a spectral density with an n pole filter, where $n = 2, 4, 6, 8$ and 16 . The relative error is defined as:

$$\text{relative error} = \left| \frac{\text{measured value} - S_{\phi}(f)}{S_{\phi}(f)} \right|$$

The potential errors increase very rapidly as the slope of $S_{\phi}(f)$ increases and as f approaches the analyzer noise bandwidth. For example, a situation which could occur would be the measurement of a high level quartz oscillator at 1 Hz . In this case the spectral density may vary as f^{-5} . If one were to make this measurement with a commercial analyzer having a noise bandwidth of 1 Hz and an 8 pole bandpass filter, then the expected error is 15 dB at $f = 2 \text{ Hz}$. Since the error depends on Fourier frequency, not only is the absolute value of the measured spectral density in error, but its measured dependence on frequency is also incorrect. It is therefore possible to draw incorrect conclusions about the fundamental noise processes which are involved. By scaling figures 1, 2 and 3 according to the noise bandwidth one may use them to estimate the minimum frequency at which $S_{\phi}(f)$ may be measured with less than a selected error limit. The accuracy of this estimate will depend on the deviations from the assumptions which have been described above.

One can draw several conclusions about how to make more accurate measurements at very low Fourier frequencies. In the first place, one should use the narrowest filters with the steepest skirt functions. Fast Fourier transform analyzers are now available which have bandwidths of a few millihertz and correspondingly steep skirts. In the second place, it is possible to lessen the severity of this problem by prewhitening the noise. One straightforward method for doing this is to use a tight phase-lock loop instead of a loose lock loop. The noise voltage which one meas-

ures, inside the loop bandwidth, is proportional to $S_y(f)$ and diverges a factor of f^2 more slowly than $S_{\phi}(f)$. The disadvantage of this technique is that the voltage to frequency coefficient of the oscillator tuning element must be measured. This coefficient may depend on Fourier frequency [2].

DOUBLE-BALANCED MIXER PHASE SENSITIVITY

In order to relate the voltage measured at the output of a mixer to the spectrum of the phase noise, it is necessary to know the phase sensitivity of the mixer (V/rad). Two significant problems are often encountered in its measurement. If two oscillators are being compared, then the mixer can be calibrated by allowing the oscillators to free run. However, very stable oscillators are often tunable over only a few Hz . One must answer the question: How can the calibration at low frequencies be extrapolated to ten or one-hundred kHz ? If two components are being compared using a phase bridge, then the possibility of obtaining a beat signal without significantly altering the measurement conditions does not exist. The calibration is often performed by inserting a signal from a second oscillator in one side of the bridge, duplicating as far as possible power level and impedance conditions. We must answer the question: How sensitive is the calibration to changes in drive level and driving impedance?

Careful measurements reveal that the answers to these questions are rather complex. Only under certain termination conditions do the mixers appear broadband, to 100 kHz , and these same conditions seriously degrade the noise floor of the measurement system. On the other hand, when the noise floor is optimized by reactively terminating the mixer, the bandwidth degrades severely and it is necessary to calibrate the system at every frequency for which phase noise measurements are desired. The results of a variety of measurements are presented in this section to demonstrate the dependence of double-balanced mixers on drive level, input impedance, and output impedance. In the following section several measurement systems are discussed which bring these variables under control and permit fast, convenient and highly accurate phase noise measurements to be made with double balanced mixers.

The measurement system which was used to characterize the double-balanced mixer is shown in Figure 4. The drive conditions are maintained nearly identical at the R and L ports of the mixer by the two isolation amplifiers which have output impedance Z_d . The mixer is terminated by an impedance Z_L and isolated from the

other loads by the LC filter. The synthesizer, in the lower signal path, makes possible beat frequency measurements of the phase sensitivity at any offset frequency. The calibrated electronic phase shifter, in the upper signal path, allows direct measurement of the phase sensitivity at any Fourier frequency by the insertion of a coherent phase modulation spectrum. It is described in more detail in the next section. For these measurements, the delay line phase-shifter is used to maintain the two input signals to the mixer in the quadrature condition. When using the same voltmeter for all measurements, the two methods agree to within 0.2 dB, so the electronic phase shifter was used for the mixer calibrations since it is significantly easier to use.

Figure 5 shows three sets of phase sensitivity calibrations. Between the two curves of each pair, the only change is the drive level which is defined as the amount of power delivered when the mixer is replaced by a 50Ω load. Both the driving impedance and the load impedance vary between the sets. These measurements reconfirm the well-known result that, independent of other conditions, the phase sensitivity depends strongly on how much current the diodes in the mixer conduct. The dependence on drive power makes it impossible to make any substitution of devices on the input of the mixer without introducing about 3 dB uncertainty in the calibration unless some method, such as the one discussed later, is used to significantly reduce this dependence.

If the mixer is terminated at its X-port by a 50Ω load, then the beat frequency output is nearly sinusoidal. When a high impedance termination, for example 1 kΩ, is used, the beat waveform becomes triangular and when the appropriate capacitive termination is used the waveform is more nearly square. This means that the slope through zero, i.e., the phase sensitivity, is increased [3]. Quantitative results are shown in Figure 6. It is seen that the available increase in sensitivity is more than 10 dB. In addition to decreasing the phase sensitivity, a 50Ω termination increases the white noise level. As a result, noise floor considerations often require the use of the capacitive termination.

The trade-off is an increased variation of the calibration with Fourier frequency. The full extent of this dependence is illustrated in Figure 7. When the driving impedance is low, 10Ω, and the load capacitance is 0.01 μF, the peak-to-peak variation in phase sensitivity is only 0.4 dB, but in the case of the constant

current drive and a 0.1 μF load the variation in phase sensitivity is 15 dB over the 1 to 100 kHz range. Neither the phase sensitivity nor its frequency dependence remain constant for even the smallest changes in a measurement system. For this reason, the figures presented here should be used only as a guide. A careful calibration must be made after every change in a measurement system. Some ways for making such calibrations more convenient are discussed in the next section.

HIGH PRECISION MEASUREMENT SYSTEMS

The following measurement systems can be calibrated to an accuracy of 0.2 dB. Differences in noise performance, convenience and means of calibration are discussed. Figure 8 illustrates a simple test set for measuring $S_{\phi}(f)$ of an oscillator. The mixer noise is relatively large and the phase sensitivity minimum. However, the phase sensitivity is flat vs. Fourier frequency. This system can be used for a wide variety of carrier frequencies with a $S_{\phi}(f)$ noise floor approximately -160 dB relative to 1 rad²/Hz. By simply measuring K_d , the phase slope at the zero crossing, using a strip chart recording or a scope trace, one can calibrate the system from dc to 100 kHz with an accuracy of about 0.2 dB. Specifically:

$$S_{\phi}(f) = \left[\frac{V_n(f)}{K_d(\text{volts/rad}) G_1 G_2} \right]^2 \frac{1}{f_h}$$

where $V_n(f)$ is the RMS noise measured with the spectrum analyzer at Fourier frequency f in a noise bandwidth f_h , and $G_1 G_2$ is the measurement system gain. It is imperative that the measurements be made with exactly the same cables and mixer drive levels as was used during the calibration. Simply replacing a cable can change the calibration factor, i.e., phase sensitivity, by 3 dB or more. Because of this, this system is not very suited for measuring noise in amplifiers, etc.

The errors associated with the spectrum analyzer bandwidth and skirt selectivity can be estimated from Figures 1, 2 and 3, if the asymptotic form of the analyzer filter roll-off is known. However, to assure an accuracy of better than 10 dB for Fourier frequencies near the filter noise bandwidth, it is necessary to map the analyzer transfer function carefully and deconvolute the measured output. A word of caution - spectrum analyzer noise bandwidths may be as much as 3 dB different than advertised. Another common trouble is due to finite dynamic range; often it is necessary to precede the spectrum analyzer with a filter to remove

out of band noise e.g., 60 Hz and its harmonics.

Another common trouble, which causes a change in mixer sensitivity, is phase drift between the oscillators due to finite gain in the phase-lock loop. This is easily avoided by using a second order phase lock loop which normally holds the phase error to less than 0.1 rad [3]. The phase lock loop attack time, τ_v , is adjusted to be equal or greater than $2/\pi f_{\min}$ where f_{\min} is the minimum Fourier frequency of interest. For Fourier frequencies below 0.1 Hz it is often necessary to measure $S_y(f)$ instead of $S_\phi(f)$. This can be accomplished by using a tight phase lock loop with τ_v less than $1/10\pi f_{\max}$, where f_{\max} is the highest Fourier frequency of interest. In this case the mixer output is proportional to $S_y(f)$.

$$S_y(f) = \frac{K_y^2}{v^2} \left[\frac{V_n}{G_1 G_2} \right]^2 \frac{1}{f_h} \quad \text{and}$$

$$S_\phi(f) = \frac{v^2}{f} S_y(f),$$

where K_y is the voltage to frequency sensitivity in Hz/volt of the oscillator tuning element, and v is the carrier frequency. If $S_\phi(f)$ is varying rapidly, the tight phase lock loop method can be used to prewhiten the noise, enabling more accurate measurements.

Figure 9 illustrates a more general measurement system which is capable of measuring $S_\phi(f)$ of signal processing equipment and passive components, as well as oscillators. The calibration procedure is very simple and simultaneously calibrates mixer sensitivity, amplifier gain and spectrum analyzer gain. This system is the most convenient one of the three for measuring $S_y(f)$ with a tight phase-lock loop.

The calibration procedure consists of driving the precision phase shifter with a reference signal of known amplitude from 100 Hz to 100 kHz. This reference mixer output is then compared to the measured noise level. The calibrated phase shifter and its calibration are discussed below. This system allows the use of a mixer termination which maximizes its phase sensitivity and minimizes the noise. For a loose phase-lock system:

$$S_\phi(f) = \left[\frac{V_n(f)}{V_R(f)} K \right]^2 \frac{1}{f_h}$$

where K is the RMS phase modulation impressed by the phase shifter in radians, and $V_R(f)$ is the output reference voltage as measured on the spectrum analyzer

as a function of f . Results obtained with this method and the beat frequency method agree to within 0.2 dB. The phase shifter can be calibrated to within 0.2 dB at dc using a time interval measurement, e.g., the dual mixer time difference system [2] or it can be calibrated with the beat frequency method to the same accuracy. At high Fourier frequencies, a storage scope can be used to check the flatness of the sensitivity. The present design, shown in Figure 10, uses a series tuned LCR circuit sandwiched between two high quality isolation amplifiers [3]. The isolation amplifiers guarantee that the phase shift is independent of input and/or output loading conditions. The LCR circuit has a Q factor of about 2 and therefore is very wideband. The residual amplitude modulation is typically -60 dB relative to the phase modulation of 1×10^{-3} rad. Measurements of $S_y(f)$ using the tight phase lock loop are calibrated using the calibrated phase shifter via:

$$S_y = \left[\frac{V_n(f)}{V_R(f)} \frac{f K}{v} \right]^2 \frac{1}{f_h},$$

where f , K , $V_n(f)$ and $V_R(f)$ have the previous definitions.

Figure 11 shows a measurement system for $S_\phi(f)$ which holds its calibration factor over a 15 dB change in input power, has very low noise, and can be made very flat from dc to 100 kHz Fourier frequency offset from the carrier. This system is based on the use of low noise current coupled pairs as odd-order multipliers [4]. This has many advantages:

- (1) The output level and drive impedance is virtually independent of input drive over a large range (See Figure 12) with the consequence that the mixer sensitivity is constant. Therefore, only one calibration is necessary.
- (2) The multiplication raises the input noise by n^2 so that the mixer noise and the following audio amplifier noise is less important. Therefore the mixer can be resistively terminated in order to keep it flat vs. Fourier frequency without sacrificing noise performance. Both oscillators and passive components can be measured.
- (3) One can freely substitute components and/or devices on the input and still maintain the same system calibration factor to within 0.1 dB, allowing precise comparisons of the added noise of various system subassemblies.
- (4) Noise floors below $S_\phi(f) = -180$ dB relative to 1 rad²/Hz have been achieved with a fundamental carrier frequency of 5 MHz and a x9 commercially available multiplier.

CONCLUSION

It has been shown that a failure to take into account the effect of the analyzer window shape and/or the frequency dependence of the mixer can lead to

errors in the measurement of $S_{\phi}(f)$ as large as 15 dB. To solve this problem, several measurement systems are presented along with calibration procedures which allow fast and convenient spectral density measurements with 0.2 dB accuracy. Although the basic concepts of these systems are already well known, their applicability to high accuracy measurements has not been demonstrated previously. To obtain an accuracy of 0.2 dB, careful calibration of the entire measurement system is an absolute requirement; with few exceptions, the most trivial change requires recalibration. The ease of measurement, calibration, and wide applicability has been stressed in the measurement systems described. One measurement system presented retains its calibration factor over a 15 dB change in input power level and possesses an inherent noise floor which is below virtually all other present signal processing equipment and oscillators.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Stephen Jarvis for calculating the errors made in estimating the spectral density as a function of the filter shape.

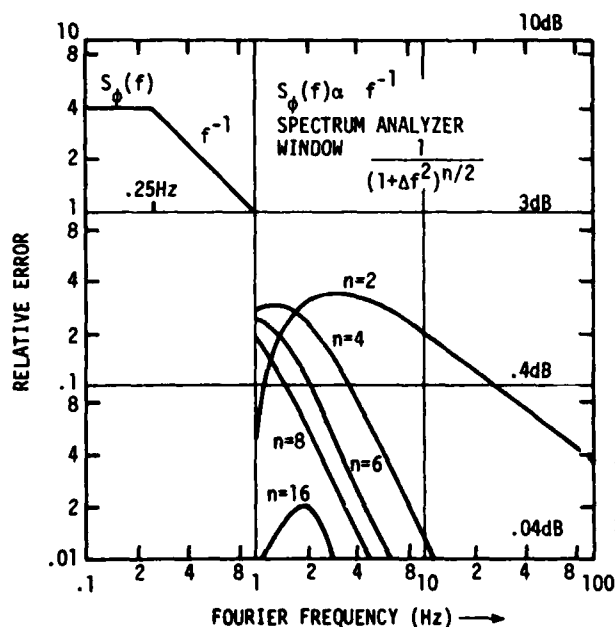


Fig. 1

The relative error of a spectral density of phase measurement as a function of Fourier frequency. The spectral density varies like f^{-1} for frequencies above 0.25 Hz as shown in the upper left hand corner. The different curves are parametrized by n , the number of poles in the bandpass filter.

REFERENCES

- [1] L.S. Cutler and C.L. Searle, "Some aspects of the theory and measurement of frequency fluctuations in frequency standards," Proc. IEEE 54, p. 136 (1966).
- [2] David W. Allan, "Measurement of frequency and frequency stability of precision oscillators," NBS Tech. Note 669, 1975.
- [3] F.L. Walls, S.R. Stein, James E. Gray, David J. Glaze and David W. Allan, "Design considerations in state-of-the-art signal processing and noise measurement systems," Proc. 30th Annual Symposium on Frequency Control, p. 269, 1976.
- [4] Richard A. Baugh, "Low noise frequency multiplication," Proc. 26th Annual Symposium on Frequency Control, p. 50, 1972.

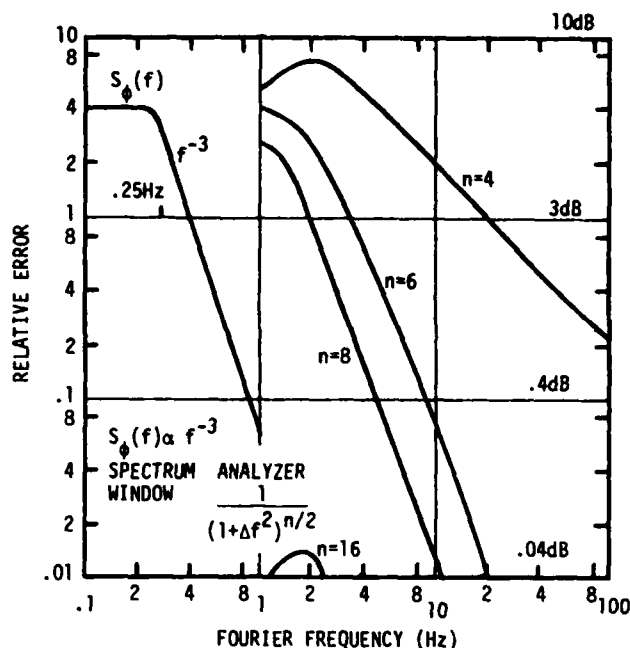


Fig. 2

The relative error of a spectral density of phase measurement as a function of Fourier frequency. The spectral density varies like f^{-3} for frequencies above 0.25 Hz as shown in the upper left hand corner. The different curves are parametrized by n , the number of poles in the bandpass filter.

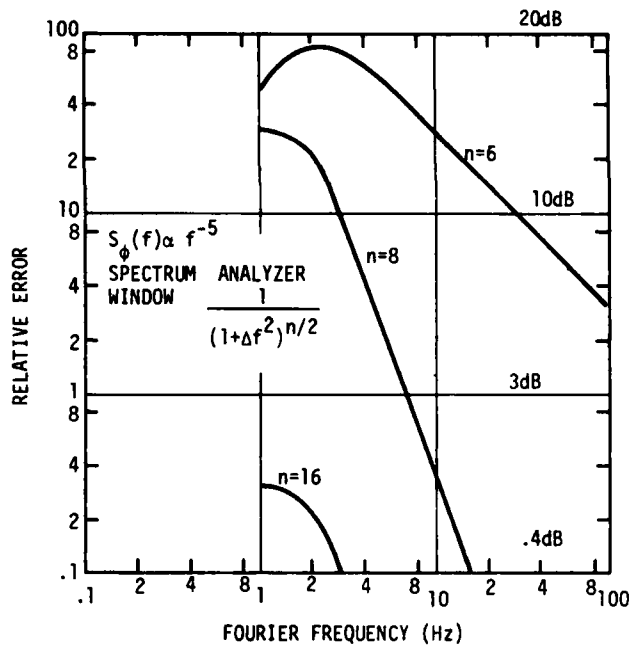


Fig. 3

The relative error of a spectral density of phase measurement as a function of Fourier frequency. The spectral density varies like f^{-5} for frequencies above 0.25 Hz. The different curves are parametrized by n , the number of poles in the bandpass filter.

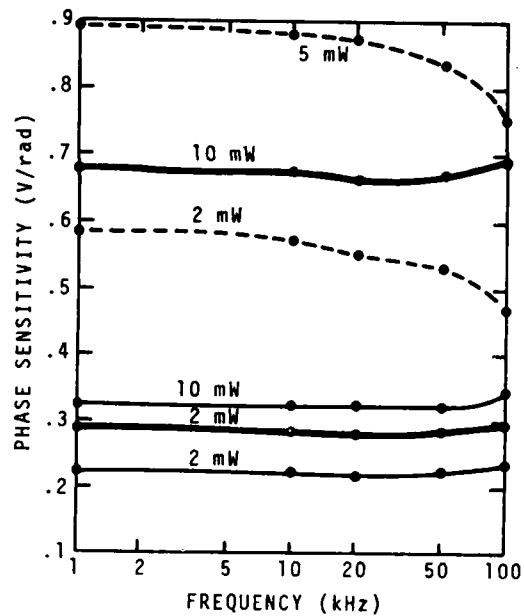


Fig. 5

The double-balanced mixer phase sensitivity as a function of Fourier frequency for different drive levels. Each pair of curves represents identical conditions except for the drive level while input/output conditions vary widely between the pairs.

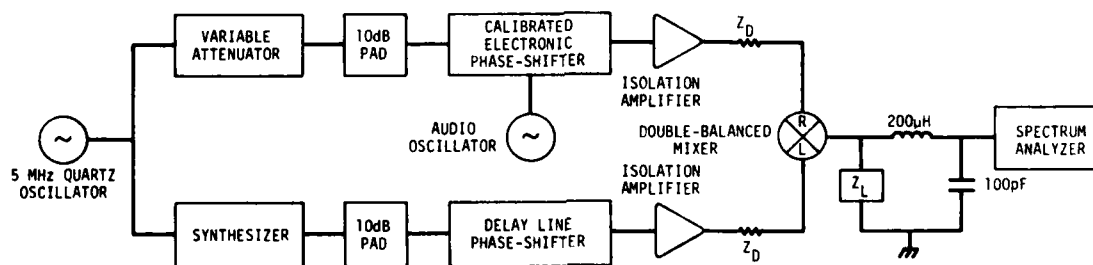


Fig. 4 Schematic of the measurement system for evaluating the dependence of the mixer phase sensitivity on drive level, drive impedance, and output impedance.

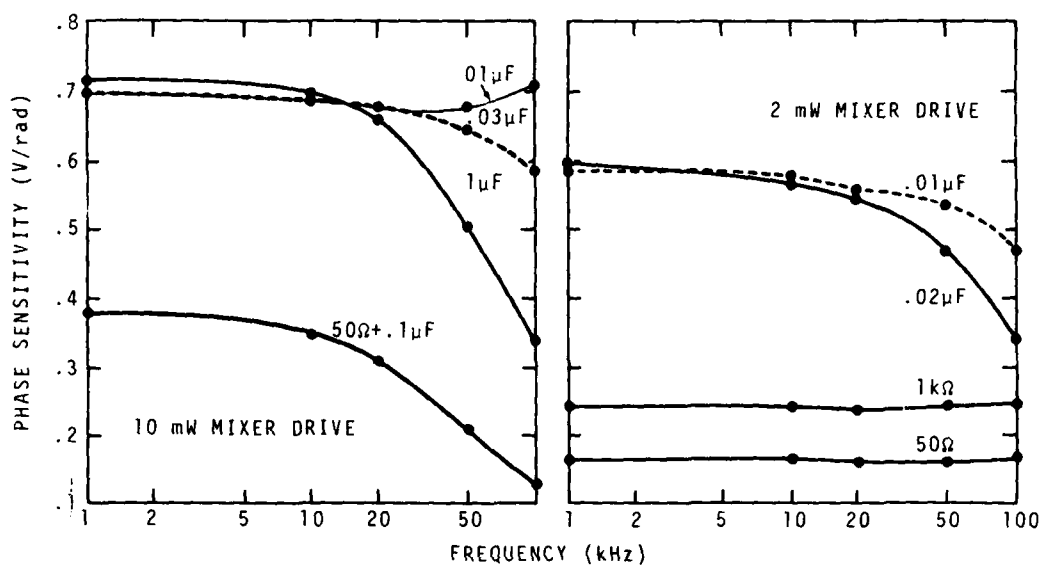


Fig. 6 Double-balanced mixer phase sensitivity as a function of frequency for various output terminations. The curves on the left were obtained with 10 mW drive while those on the right were obtained with 2 mW drive. The data demonstrate a clear choice between constant, but low sensitivity or much higher, but frequency dependent sensitivity.

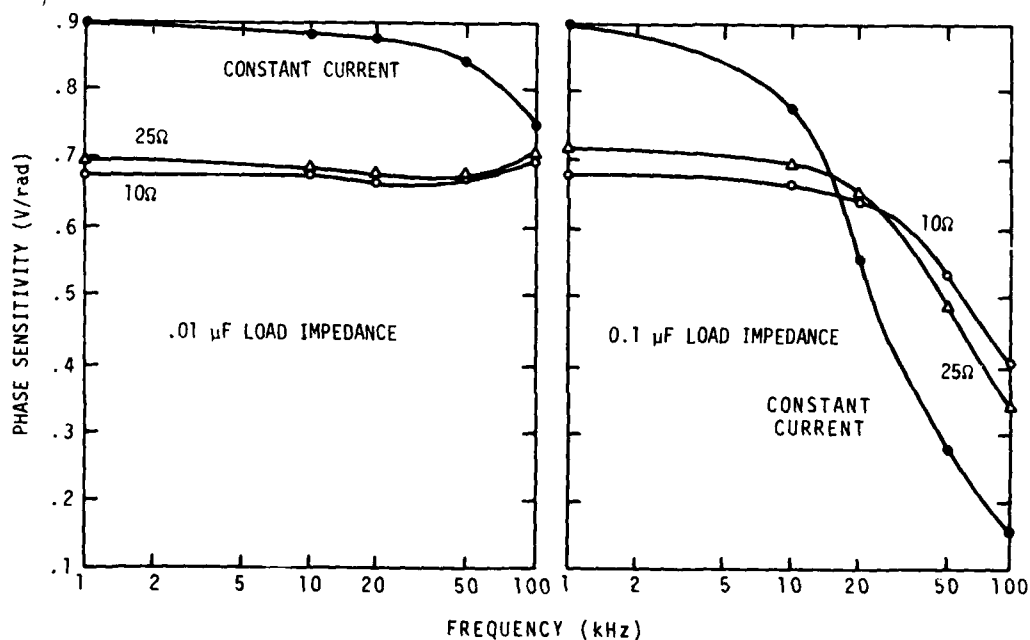


Fig. 7 Double-balanced mixer phase sensitivity as a function of Fourier frequency for different driving impedances. The curves on the left and right differ in the value of the phase output termination. The frequency dependence of the phase sensitivity depends strongly on the value of the termination capacitor.

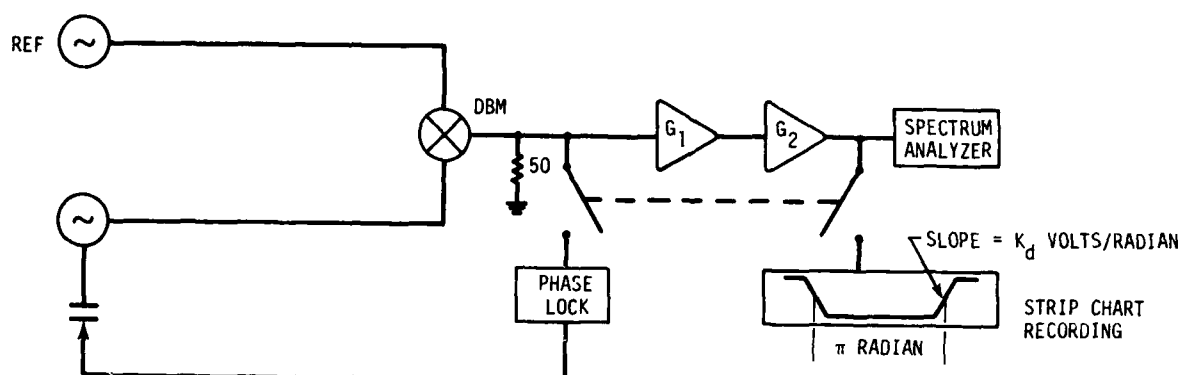


Fig. 8 Precision phase measurement system. Calibration requires an external strip chart or storage scope to measure the slope at the zero crossing. The accuracy is 0.2 dB from dc to 100 kHz Fourier frequency offset from the carrier. This system is suitable only for oscillators. A wide range of carrier frequencies can be accommodated with the same measurement system.

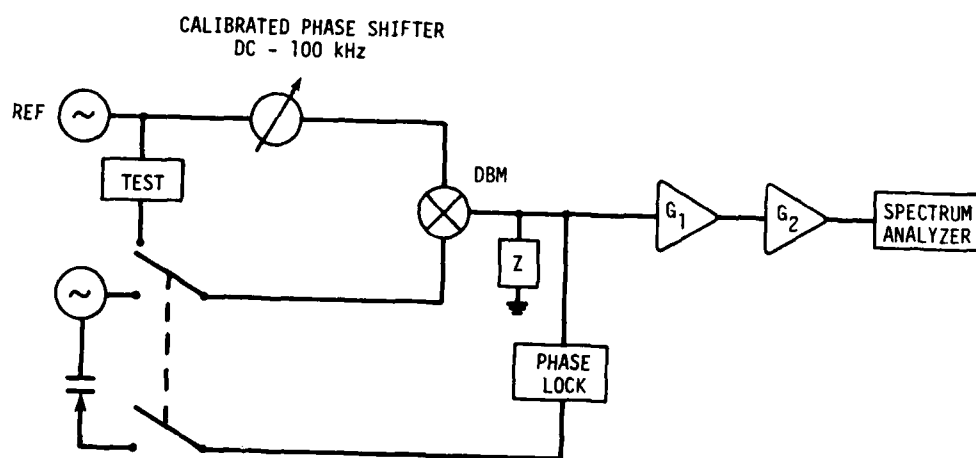


Fig. 9 Precision phase measurement system featuring self calibration to 0.2 dB accuracy from dc to 100 kHz Fourier frequency offset from carrier. This system is suitable for measuring signal handling equipment and passive components, as well as oscillators. Different carrier frequencies require a different phase shifter.

CALIBRATED PHASE SHIFTER

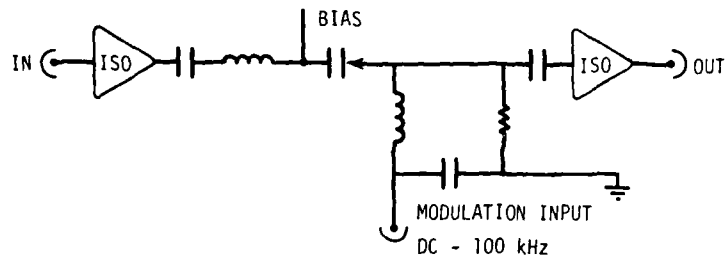


Fig. 10 Isolated phase modulator is insensitive to input or output loading. The LCR series tuned circuit has a Q of about 2. The phase modulation sensitivity is constant to 0.1 dB from dc to 100 kHz.

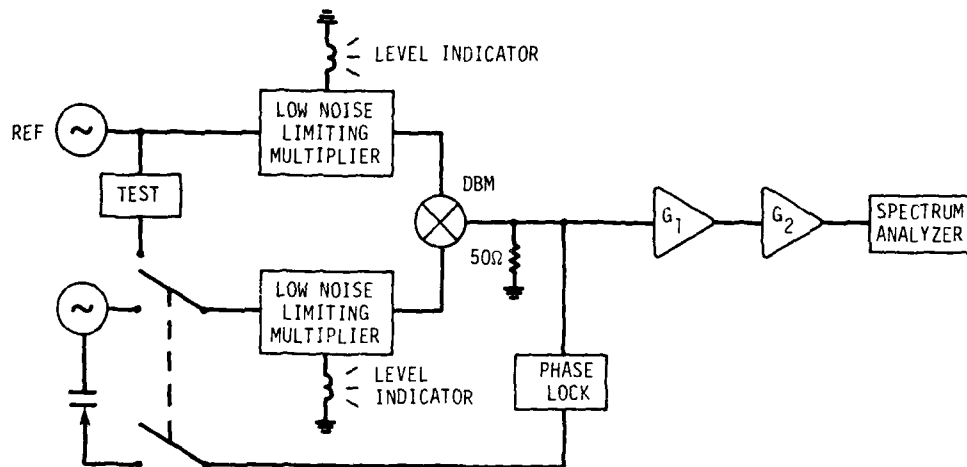


Fig. 11 Precision phase measurement system featuring constant calibration factor for a wide range of input signal levels and Fourier frequency. This system also features the lowest noise floor, about -180 dB relative to 1 rad/Hz and is the most suitable for measuring phase noise of signal handling devices, passive components or oscillators. Different carrier frequencies require a different pair of multipliers.

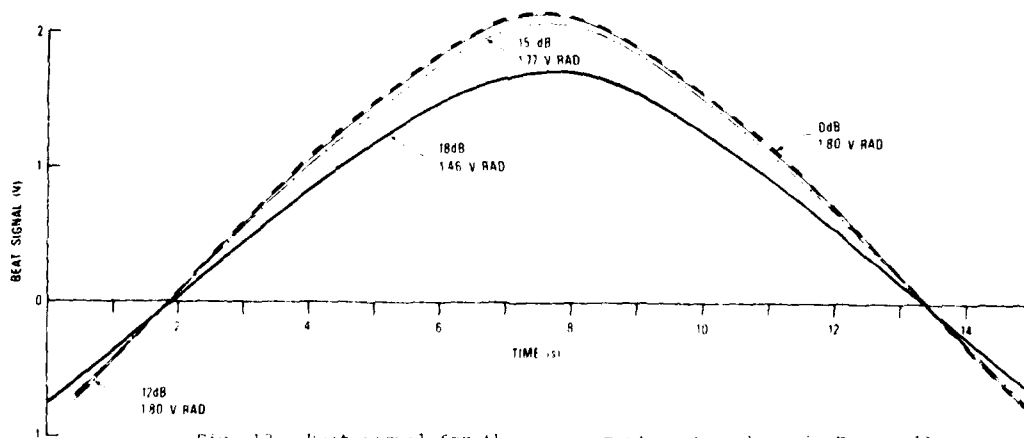


Fig. 12 Beat signal for the measurement system shown in Figure 11 as a function of attenuation of the input drive level.

FREQUENCY AND TIME DOMAIN STABILITY OF THE Rb^{87} MASER AND RELATED OSCILLATORS: A PROGRESS REPORT

Jacques Vanier, Michel Têtu, Roger Brousseau

Laboratoire d'Electronique Quantique
Département de Génie Electrique
Université Laval

Abstract

Time domain and frequency domain characteristics of the rubidium maser are reported. The two types of measurements are found to be in relatively good agreement. The measurements show that the maser is one of the oscillators with the highest spectral purity and which consequently has a very good short term stability. The question of transferring short term stability from a good oscillator to a poorer oscillator is examined.

Introduction

In the past we have given reports on the construction of a small rubidium maser (1) as well as on its physical characteristics (2). The maser itself is a unit 20 cm in diameter by 30 cm long. Its complete description is given in reference (1). Its stability has been reported in the time domain as $2 \times 10^{-13} \tau^{-1}$ for $\tau < 1$ sec; a flicker floor was obtained at a level of about $2 \times 10^{-13} \tau^0$ for $1 \text{ sec} < \tau < 10^3$. The measurements in the short term region were done on the beat frequency between two rubidium masers and in the mid term region between the rubidium maser and a hydrogen maser.

More recently we have made measurements on the frequency stability of this type of maser in the frequency domain and have related these measurements to results obtained in the time domain. We have examined the problem of the transfer of stability in the short term region, from the maser to a quartz oscillator of poorer stability. This paper reports some progress on both of these subjects.

Rubidium Maser Frequency Stability

A typical result of a measurement of the rubidium maser stability obtained in the time domain is shown in fig. 1. These measurements were made on the beat signal between the compact Rb^{87} maser described in reference (1) and another Rb^{87} maser (3), the first maser being considered as the measurement reference. The results shown in fig. 1 have not been divided by $\sqrt{2}$. Two regions are clearly identified. A region where $\sigma(\tau)$ varies as τ^{-1} which is considered as due to white phase noise. The source of this noise is believed to be thermal noise added to the maser signal by the receiver. In a second region $\sigma(\tau)$ varies as $\tau^{-1/2}$ which reflects a random walk type of frequency instability. Here this instability is believed to originate mostly from the thermal fluctuations of the maser cavity frequency through the pulling phenomenon.

The frequency domain measurements were done with the set up shown in fig. 2. In that case a low frequency stable oscillator was used as the reference and was phase locked to the beat signal between the two masers. The system could be operated either in the tight phase lock mode or in the loose phase lock mode. The spectrum analyser used was a Schlumberger EMR model no 1510. The limit of resolution of the system was about 20 db below the lowest noise level measured. The results of the measurements are shown in fig. 3 as $S_y(f)$: the spectral density the fractional frequency instability y , as a function of the Fourier frequency. Two

regions are again clearly identified and the behaviour of the masers can be represented by the expression:

$$S_y(f) = 1.5 \times 10^{-25} f^{-2} + 3.6 \times 10^{-27} f^2$$

The first term represents a random walk type of frequency instabilities and the second term, a white phase noise contribution. These measurements can be expressed in the time domain through the usual transformation formula (4). This gives:

$$\tau > 1 \text{ sec} \rightarrow \sigma_y(\tau) = 1 \times 10^{-12} \tau^{\frac{1}{2}}$$

(random walk of frequency).

$$\tau < 1 \text{ sec} \rightarrow \sigma_y(\tau) = 1.6 \times 10^{-14} f_h^{\frac{1}{2}} \tau^{-1}$$

(white phase noise) where f_h is the cut off frequency of the filter used in the measuring set up. It was verified that most of the white phase noise originated from additive noise probably generated in the receiver. In fact a wave analyser connected directly at the output of the receiver gave a noise level expressed in the same unit as the phase spectral density of $S_y(f)$ of -73.5 db relative to one rad^2 per Hz. This corresponds to a stability $\sigma(\tau) = 8.5 \times 10^{-15} f_h \tau^{-1}$ which is close to the results given above. A similar measurement made on the output of a hydrogen maser receiver equipped with a pre-amplifier at 1420 MHz and 2 db noise figure gave an equivalent white phase noise limit, S_y of -54 db relative to one rad^2 per Hertz. This gives a $\sigma(\tau)$ of $3.9 \times 10^{-13} f_h \tau^{-1}$ in the short term region, consistent with published results in such devices. Thermal noise within the maser line width will also set a lower limit to the stability of these devices in the short term observation times. In the case of the hydrogen maser one obtains for an optimum setting of all parameters $\sigma(\tau) = 2 \times 10^{-14} \tau^{-1/2}$ while for the rubidium maser one obtains $\sigma_y(\tau) = 4 \times 10^{-14} \tau^{-1/2}$ (5).

Crystal Oscillator Phase Locked to the Maser

The atomic frequency standard has in general an output frequency which is not readily usable. It is then standard practice to phase lock a crystal oscillator at a standard frequency of, say, 10 MHz to the atomic standard. This is shown in fig. 4 where an integrator is placed in the feedback loop. At the condition of critical damping, the spectral density of the phase fluctuations of the locked crystal oscillator are given by:

$$S_{\phi_l}(\omega) = \frac{(\omega \tau')^4}{(1 + \omega \tau')^2} S_{\phi_r}^f(\omega) + \frac{1}{n^2} \frac{1 + 4(\tau')^2}{(1 + \omega \tau')^2} S_{\phi_r}(\omega) \quad (1)$$

where S_{ϕ_l} is the phase spectral density of the crystal oscillator, S_{ϕ_r} is the spectral density of the reference, ω is a Fourier frequency, l and r refer respectively to locked and free running mode and τ' is a time constant defining loosely the bandwidth of the phase lock loop. As shown in the same figure, the first term acts as a high pass filter. In this way the low frequency fluctuations of the crystal oscillator are filtered out or corrected. The phase stability of the

maser is thus transferred to the crystal oscillator for low Fourier frequencies or in other words the long term stability of the maser is transferred to the crystal oscillator. One question arises, however, regarding the stability of the crystal oscillator in the short term region. It has been standard practice to use crystal oscillators with very low phase noise at high Fourier frequencies and adjust the bandwidth of the feedback filter to a value that optimizes the performance of the whole frequency standard. The possibility of using low quality crystal oscillators (high level of white phase noise) would be interesting. In that case the bandwidth of the phase lock loop would need to be opened in order for phase fluctuations at high Fourier frequencies to be filtered, as is shown in fig. 4.

Experimental attempts have been made in order to check qualitatively these possibilities. The reference oscillator was a good crystal oscillator at 5 MHz with white phase noise below -130 db relative to one rad² per Hertz. The quartz oscillator was a simple device made of a quartz crystal at 5 MHz coupling the input and output of two cascaded invertors. In the present situation the multiplication factor n in equation (1) is unity. Fig. 5 gives the results of S_{ϕ} in the case the bandwidth of the phase lock loop f' is 1 Hz. The curves identified as 10 μ V and 45 μ V correspond to cases where the oscillator tuning varactor was modulated with a noise generator such as to create white phase noise over a broad range of Fourier frequencies. Fig. 6 and 7 show the results obtained when the bandwidth of the phase lock loop was opened corresponding to respective situations where $f' = 110$ Hz and $f' = 276$ Hz. In this case it is clearly observed that part of the phase noise is filtered in a way compatible with the bandwidth set and in qualitative agreement with equation 1. The fact that the noise at the mixer does not drop to the same level at low frequencies for both amplitudes of noise applied to the crystal oscillator may be due to the finite gain of the operational amplifier used in the filter of the phase lock loop. From these results it is clear that measurements of the "noisy" quartz oscillator stability in the time domain would be improved. However care would need to be exercised in adjusting the bandwidth of the time domain measuring equipment to a value compatible with that of the phase lock loop system. Furthermore the slope of the filter used in the time domain measurement would need to be at least 40 db per decade, otherwise noise at high Fourier frequencies would predominate and mask the improvement observed at low Fourier frequencies.

Conclusion

In the present paper we have reported the results of measurements of frequency stability in both time and frequency domains. We have found that the two types of measurements are consistent in the present case, and have identified the various types of noise processes contribution to the instabilities measured. From these measurements and past studies it appears that the rubidium maser could be largely improved by control of proper parameters such as light shift and temperature. These are believed to be the factors affecting the medium term stability through a random walk type of noise process. As far as white phase noise is concerned, the maser still appears as one of the best oscillators that can be used as a reference oscillator.

From the preliminary results on the phase lock loop question it appears that in principle, white phase noise at the output of a crystal oscillator can be reduced to some extent by a proper choice of the bandwidth of the loop. This indicates that the short term

stability of a good low phase noise oscillator such as a maser could, in principle, be transferred to a crystal oscillator of poorer short term stability. However further experiments are required to verify that the actual technique used in the experiment described for introducing white phase noise in the phase locked crystal oscillator is truly equivalent to the use of a poor crystal oscillator having a high level of white phase noise.

Acknowledgement

The authors would like to thank Dr Claude Audoin for his suggestions relative to the phase lock loop problem and Mr. R. Blier for his competent technical help. One of us (Jacques Vanier) would like to thank the personnel of the Time and Frequency Division of the National Bureau of Standards where he has spent a period of time and has benefited from very stimulating discussions.

References

- (1) A. Compact Rb⁸⁷ Maser, G. Busca, J. Racine, J. Vanier, Prog. 28th Annual Symposium on Frequency Control, 1974, p. 344.
- (2) Long Term Frequency Stability of the Rb⁸⁷ Maser, G. Busca, R. Brousseau, J. Vanier, IEEE Trans. on Instr. and Meas., IM 24, 1975, p. 291.
- (3) Short Term Frequency Stability of the Rb⁸⁷ Maser, M. Têtu, G. Busca, J. Vanier, IEEE Trans. on Instr. Meas., IM 22, 1973, p. 250.
- (4) Characterization of Frequency Stability, J.A. Barnes et al., IEEE Trans. on Instr. and Meas., IM 20, 1971, p. 105.
- (5) On the Fundamental Limitation of Masers Short Term Frequency Stability, G. Busca, S. Rovea, to be published.

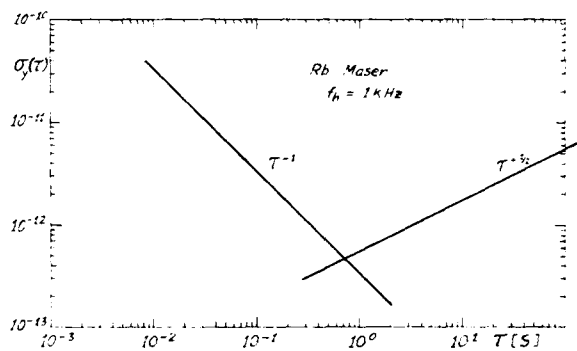


FIGURE 1. Time domain frequency stability of the rubidium maser; $\sigma_v(\tau)$ is the square root of the sample variance.

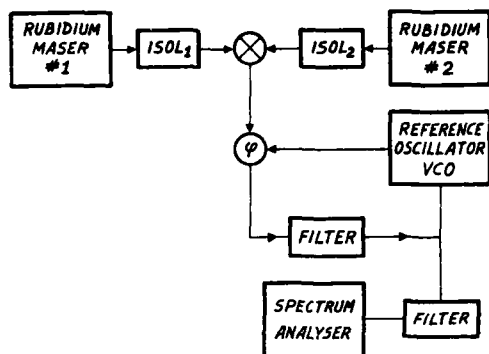


FIGURE 2. System used to measure the frequency stability of a pair of rubidium masers in the frequency domain.

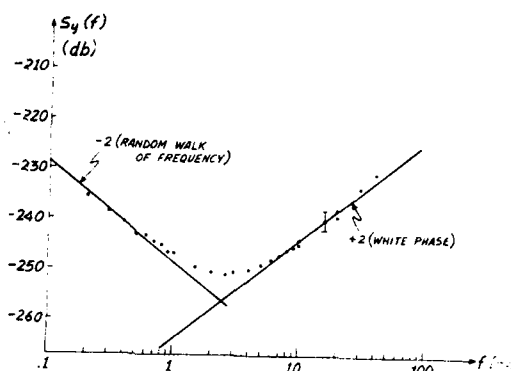
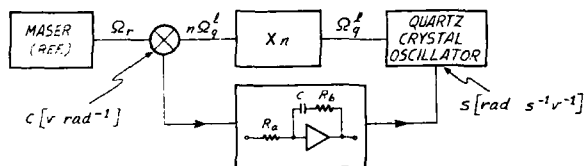


FIGURE 3. Frequency stability of two rubidium masers as measured in the frequency domain.



$$T_a = CR_a \quad T_b = CR_b \quad \omega_n = \sqrt{\frac{n^2 C}{T_a}} \quad \zeta = \frac{\omega_n T_b}{2}$$

$$\text{CRITICAL DAMPING: } T' = \frac{T_b}{2} \quad T_b = 4TT_a \quad T = (n\zeta C)^{-1}$$

$$\Rightarrow S_{\phi_q}(\omega) = \left\{ \frac{(\omega T')^4}{[1 + (\omega T')^2]^2} \right\} S_{\phi_r}(\omega) + \frac{1}{n^2} \left\{ \frac{1 + 4(T'\omega)^2}{[1 + (\omega T')^2]^2} \right\} S_{\phi_r}(\omega)$$

FIGURE 4. Diagram of a typical phase lock loop used to lock a crystal oscillator to a maser.

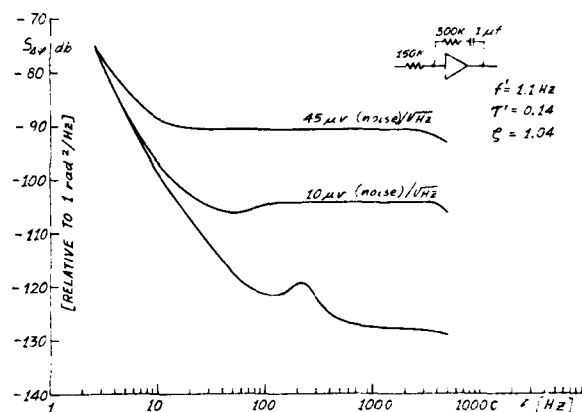


FIGURE 5. Spectral density of phase as a function of the Fourier frequency for a "noisy" quartz crystal oscillator as measured with a "loose phase lock loop".

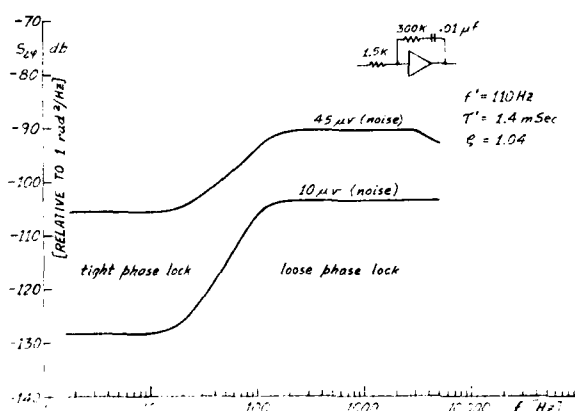


FIGURE 6. Effect of increasing the bandwidth of the phase lock loop on the spectral density of phase of the "noisy" quartz crystal oscillator. The bandwidth of the servo loop is 110 Hz.

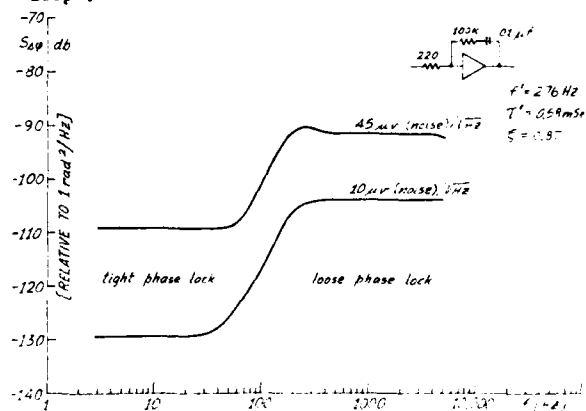


FIGURE 7. Effect of increasing the bandwidth of the phase lock loop on the spectral density of phase of the "noisy" quartz crystal oscillator. The bandwidth of servo loop is 276 Hz.

AUTOMATING PHASE NOISE MEASUREMENTS IN THE FREQUENCY DOMAIN

Algie L. Lance, Wendell D. Seal,
Frank G. Mendoza and Neal W. Hudson

TRW Defense and Space Systems Group
Equipment Management Center, Metrology
One Space Park
Redondo Beach, California 90278

Summary

Automation of phase noise measurements has been developed with satisfactory results using two types of measurement systems. Measurements are performed in the frequency domain using a spectrum analyzer which provides a frequency window following the phase or frequency detector. One system measures the combined phase noise characteristics of two sources. The two source signals are applied in quadrature to a phase sensitive detector (double balanced mixer) and the voltage fluctuations analog to the phase fluctuations are measured at the detector output. One measurement system is designed to measure the phase noise characteristics of a single oscillator. The single-oscillator measurement system is designed using the delay line as an FM discriminator. Voltage fluctuations analog to frequency fluctuations are measured at the detector output.

The calibration and measurement steps are controlled by a calculator program. The calibration sequence requires several manual operations. The software program controls frequency selection, bandwidth settings, settling time, amplitude ranging, measurements, calculations, graphics, and data plotting.

A quasi-continuous plot of phase noise is obtained by performing measurements at Fourier frequencies separated by the IF bandwidth of the spectrum analyzer used during the measurement. Plots of other defined parameters can be obtained and plotted as desired. Data plots will be used to show the effects of phase lock loops, inadequate isolation of two sources when using the two-oscillator technique, and the limiting frequency range of measurements performed using the delay line as an FM discriminator.

High quality sources can be measured without multiplication to enhance the phase noise prior to downconverting and measuring at baseband frequencies.

Measurements can be extended through the millimeter frequency band. The integrated phase noise can be calculated for any selected range of Fourier frequencies. This capability is being added to the present software program.

Special reference -- Automated Phase Noise Measurements, The Microwave Journal, Vol. 20, No. 6, June, 1977.

Introduction

In this presentation the Greek letter nu (ν) represents frequency for carrier-related measures. Modulation-related frequencies are designated (f). If the carrier is considered as dc, the frequencies measured with respect to the carrier are referred to as baseband, offset from the carrier, modulation, noise, or Fourier frequencies.

The term frequency stability encompasses the concepts of random noise, intended and incidental modulation, and any other fluctuations of the output frequency of a device. In general, frequency stability is the degree to which an oscillating source produces the same value of frequency throughout a specified period of time.

Long-term stability is usually expressed in terms of parts per million per hour, day, week, month or year. This stability represents phenomenon due to the aging process of circuit elements and of the material used in the frequency determining element. Short-term stability relates to frequency changes of less than a few seconds duration about the nominal frequency. The measurement of frequency stability can be accomplished in both the time domain and frequency domain. In the frequency domain, measurements are performed using a spectrum analyzer which provides a frequency window following the detector. In the time domain, measurements are performed with a gated counter which provides a time window following the detector.

Phase noise is the term most widely used to describe the characteristic randomness of frequency. The term spectral purity refers to the signal power to phase noise sideband power ratio. Spurious signals are outputs in the spectrum of a source that are neither part of the carrier nor its harmonics, and they may be discrete frequencies or bands of frequencies. The spectral content of the noise modulating signals is finite in bandwidth and does not have the same bandwidth and same amplitude at all frequencies as one finds with white noise.

Practical oscillators demonstrate noise which appears to be a combination of casually generated signals and random, non-deterministic noises. The random noises include thermal noise, shot noise, noises of undetermined origin (such as flicker noise), and integrals of those noises. The end result is time dependent phase and amplitude fluctuations.

General Theory and Definitions

The instantaneous output voltage $V(t)$ of a signal generator or oscillator may be written as ¹

$$V(t) = [V_0 + \epsilon(t)] \sin [2\pi\nu_0 t + \phi(t)] \quad (1)$$

where V_0 and ν_0 are the nominal amplitude and frequency respectively and $\epsilon(t)$ and $\phi(t)$ are the instantaneous amplitude and phase fluctuations of the signal. The definition of instantaneous fractional frequency deviation $y(t)$ is based on the provision that $\epsilon(t)$ and $d\phi/dt$ are sufficiently small for all time t . $\phi(t)$ is a varying real function of time, representing angle modulation, i.e., frequency and phase fluctuation.

Frequency fluctuations ($\delta\nu$) are related to phase fluctuations ($\delta\phi$) by

$$\delta\nu \equiv \delta\Omega/2\pi = 1/2\pi \, d(\delta\phi)/dt \quad (2)$$

i.e., angular frequency deviation $\delta\Omega$ is equal to the rate of change of phase deviation (the first time derivative of the instantaneous phase deviation).

y is defined as the fractional frequency fluctuation or normalized frequency deviation. It is the value of $\delta\nu$ normalized to the average (nominal) signal frequency ν_0 .

$$y \equiv \delta\nu/\nu_0 \quad [\text{dimensionless}] \quad (3)$$

$y(t)$ is the instantaneous fractional frequency deviation from the nominal frequency ν_0 .

A representation of fluctuations in the frequency domain is a graph called spectral density. Spectral density is the distribution of variance versus frequency.

The spectral density $S_y(f)$ of the instantaneous fractional frequency fluctuations $y(t)$ is defined as a measure of frequency stability¹.

$S_y(f)$ is the one-sided spectral density of fractional frequency fluctuations on a "per hertz" basis; the dimensionality is Hz^{-1} .

$$S_y(f) = S_{\delta\nu}(f)/\nu_0^2 \quad [\text{Hz}^{-1}] \quad (4)$$

$S_{\delta\nu}(f)$, in Hz^2/Hz , is the one-sided spectral density of frequency fluctuations $\delta\nu$. It is calculated as $(\delta\nu_{\text{rms}})^2/(\text{Bandwidth used in the measurement of } \delta\nu_{\text{rms}})$.

The spectral density of phase fluctuations is a frequency domain measure of phase fluctuations defined as follows:

$S_{\delta\phi}(f)$, in rad^2/Hz , is the one-sided spectral distribution of the phase fluctuations on a "per hertz" basis. It is calculated as

$$S_{\delta\phi}(f) = \frac{(\delta\phi_{\text{rms}})^2}{\text{Bandwidth used to meas. } \delta\phi_{\text{rms}}} \quad [\text{rad}^2/\text{Hz}] \quad (5)$$

The phase and fractional frequency fluctuation spectral densities are related by:

$$S_{\delta\phi}(f) = (\nu_0^2/f^2) S_y(f) \quad [\text{rad}^2/\text{Hz}] \quad (6)$$

$S_{\delta\Omega}(f)$, in $(\text{rad/s})^2/\text{Hz}$, is the spectral density of angular frequency fluctuations $\delta\Omega$. The defined spectral densities have the following interconnecting relationships.

$$\begin{aligned} S_{\delta\nu}(f) &= \nu_0^2 S_y(f) = (1/2\pi)^2 S_{\delta\Omega}(f) \\ &= f^2 S_{\delta\phi}(f) \quad [\text{Hz}^2/\text{Hz}] \quad (7) \end{aligned}$$

and

$$\begin{aligned} S_y(f) &= S_{\delta\nu}(f)/\nu_0^2 = (1/2\pi)^2 (1/\nu_0)^2 S_{\delta\Omega}(f) \\ &= \frac{f^2}{\nu_0^2} S_{\delta\phi}(f) \quad [\text{Hz}^{-1}] \quad (8) \end{aligned}$$

A useful measure of frequency stability relates the sideband power associated with phase fluctuations to the carrier power level. The defined measurand is called Script $\mathcal{L}(f)$. $\mathcal{L}(f)$ is defined as the ratio of the power in one sideband, referred to the input carrier frequency, on a per hertz of bandwidth spectral density basis, to the total signal power, at Fourier frequency f from the carrier, per one device². It is a normalized frequency domain measure of phase fluctuation sideband power.

$$\mathcal{L}(f) \equiv \quad [\text{Hz}^{-1}] \quad (9)$$

Power Density (one phase modulation sideband)
Carrier Power

For the condition that the phase fluctuations occurring at rates f and faster are small compared to one radian, a good approximation is²

$$S_{\delta\phi}(f) \Big|_{\text{one device}} = (2 \text{ rad}^2) \mathcal{L}(f) \quad [\text{rad}^2/\text{Hz}] \quad (10)$$

If the small angle condition is not met, Bessel function algebra must be used to relate $S_{\delta\phi}(f)$ to $\mathcal{L}(f)$.

Script $\mathcal{L}(f)$ often is expressed in decibels relative to the carrier per hertz (dBc/Hz) which is calculated as

$$\mathcal{L}(f) = 10 \log \left[\frac{S_{\delta\phi}(f) \Big|_{\text{one unit}}}{2 \text{ rad}^2} \right] \quad [\text{dBc/Hz}] \quad (11)$$

It is very important to note that the theory, definitions and equations given above relate to the noise of a single device.

Noise Processes

The spectral density plot of a typical oscillator's output is usually a combination of different noise processes. It is very useful and meaningful to categorize these processes since the first job in evaluating a spectral density plot is to determine which type of noise exists for the particular range of Fourier frequencies. The two basic categories of noise are the discrete frequency noise and the power-law noise process.

Discrete frequency noise is a type of noise in which there is a dominant observable probability i.e., deterministic in that they can usually be related to the mean frequency, power line frequency, vibration frequencies or AC magnetic fields, or Fourier components of the nominal frequency. These frequencies can have their own spectral density plots which can be defined as noise on noise.

Power-law noise process - Types of noise which produce a certain slope on the one sided spectral density plot. They are characterized by their dependence on frequency. The spectral density plot of a typical oscillator output is usually a combination of the different power-law processes. In general, we can classify the power-law processes into five categories. These five processes are illustrated in Figure 1 which can be referred to with respect to the following description of each process².

Random Walk FM - (random walk of frequency) - The plot goes down as $1/f^4$. This noise is usually very close to the carrier and is difficult to measure. It is usually related to the oscillator's physical environment (mechanical shock, vibration, temperature, or other environmental effects).

Flicker FM - (flicker of frequency) - The plot goes down as $1/f^3$. This noise is typically related to the physical resonance mechanism of the active oscillator or the design or choice of parts used for the electronics or power supply, or even environmental properties. The time domain frequency stability over extended periods is constant. In high quality oscillators this noise may be masked by white FM ($1/f^2$) or flicker phase modulation ϕM ($1/f$). It may be masked by drift in low quality oscillators.

White FM - (white of frequency) - Random Walk of Phase - Plot goes down as $1/f^4$. A common type of noise found in passive-resonator frequency standards. Cesium and Rubidium frequency standards have white FM noise characteristics since the oscillator (usually quartz) is locked to the resonance feature of these devices. This noise gets better as a function of time until it (usually) becomes flicker FM ($1/f^3$) noise.

Flicker ϕM - (flicker (modulation) of phase) - The plot goes down as $1/f$. This noise may relate to the physical resonance mechanism in an oscillator. It is common in the highest quality oscillators. This noise can be introduced by noisy electronics, amplifiers necessary to bring the signal amplitude up to a usable level, and frequency multipliers. This noise can be reduced by careful design and hand-selecting all components.

White ϕM - (white of phase) - White phase noise plot is flat f^0 . Broadband phase noise is generally produced in the same way as flicker ϕM ($1/f$). Late stages of amplification are usually responsible. This noise can be kept low by careful selection of components and narrow band filtering at the output.

The interconnecting relationships of modulation index, rms frequency deviation, peak frequency deviation, and spectral density of phase fluctuations are shown in Figure 1. These relationships are derived in Appendix A.

Basic Two-Oscillator Technique

A functional block diagram of the measurement system employing two oscillators is shown in Figure 2. NBS has performed phase noise measurements using this basic type system since 1967.

Assume that the reference oscillator is perfect (no phase noise) and that it can be adjusted in frequency. Also, assume that both oscillators are extremely stable so that phase quadrature can be maintained without the use of a phase-lock loop.

The double balanced mixer acts as a phase-sensitive detector so that when two signals are identical in frequency and nominally are in phase quadrature, the mixer output is a small fluctuating voltage δv centered on approximately zero volts. This small fluctuating voltage represents the phase modulation PM sideband components of the signal.

If the two oscillator signals applied to the mixer of Figure 2 are slightly out of zero beat, a slow sinusoidal voltage with a peak-to-peak voltage of V_{p-p} can be measured at the mixer output. Under the conditions for which Equation (10) is valid, the relationship between mean-square fluctuations of phase $\delta\phi$ and voltage δv interpreted in a spectral density fashion has been shown to be¹

$$S_{\delta\phi}(f) = S_{\delta v}(f)/2(V_{rms})^2 \quad [\text{rad}^2/\text{Hz}] \quad (12)$$

Here, $S_{\delta v}(f)$, in volts squared per hertz, is the spectral density of the voltage fluctuations at the mixer output. Since the spectrum analyzer measures rms voltage in a bandwidth B , the analyzer reading is in units of volts per $\sqrt{\text{Bandwidth}}$. Therefore,

$$S_{\delta v}(f) = [\delta v_{rms}/\sqrt{B}]^2 = \frac{(\delta v_{rms})^2}{B} \quad [\text{V}^2/\text{Hz}] \quad (13)$$

where B is the noise power bandwidth used in the measurement.

Since it was assumed that the reference oscillator did not contribute any phase noise, the voltage fluctuations $(\delta v)_{rms}$ represent the oscillator under test and the spectral density of the phase fluctuations in terms of the voltage measurements performed with the spectrum analyzer is

$$S_{\delta\phi}(f) = \frac{1}{2} \frac{(\delta v_{rms})^2}{B(V_{rms})^2} \quad [\text{rad}^2/\text{Hz}] \quad (14)$$

Equation 12 is sometimes expressed as³

$$S_{\delta\phi}(f) = S_{\delta v}(f)/K^2 \quad [\text{rad}^2/\text{Hz}] \quad (15)$$

where K is the calibration factor in volts per radian.

For sinusoidal beat signals, the peak voltage of the signal equals the slope of the zero crossings in volts per radian. Therefore, $(V_{peak})^2 = 2(V_{rms})^2$, which is the same as the denominator in Equation 12.

$S_{\delta\phi}(f)$ can be expressed in terms of decibels relative to one square radian per hertz by calculating $10 \log S_{\delta\phi}(f)$ of the previous equation.

$S_{\delta\phi}(f)$ in decibels relative to $1 \text{ rad}^2/\text{Hz}$ =

$$[20 \log [(\delta v)_{rms}] - 20 \log(V_{rms}) - 10 \log(B) - 3 \text{ dB}] \quad [\text{dB}(\text{rad}^2/\text{Hz})] \quad (16)$$

A correction of +2.5 dB to the $(\delta v)_{rms}$ term is required for the Tracking Spectrum Analyzer used in these measurement systems. See HP Application Note 207.³

$\mathcal{L}(f)$ can be expressed in terms of decibels relative to the carrier power per hertz (dBc/Hz).

$\mathcal{L}(f)$ in decibels relative to 1 Hz^{-1} =

$$[20 \log [(\delta v)_{rms}] - 20 \log(V_{rms}) - 10 \log(B) - 6 \text{ dB}] \quad [\text{dBc/Hz}] \quad (17)$$

The -6 dB correction occurs because the operation of the mixer, when it is driven at quadrature, is such that the amplitudes of the two phase sidebands add linearly in the output of the mixer. This results in four times as much power in the output as would be present if only one of the phase sidebands were allowed to contribute to the output of the mixer.

Spectrum analyzer corrections must also be added to this equation.

Two Noisy Oscillators

The measurement system of Figure 2 yields the output noise from both oscillators. If the reference oscillator is superior in performance as assumed in the previous discussions, then one obtains a direct measure of the noise characteristics of the oscillator under test.

If the reference and test oscillators are the same type, a useful approximation is to assume that the measured noise power is twice that which is associated with one noisy oscillator. This approximation is in error by no more than 3 dB for the noisier oscillator even if one oscillator is the major source of noise. The equation for the spectral density of measured phase fluctuations is

$$S_{\delta\phi}(f) \Big|_{\#1} + S_{\delta\phi}(f) \Big|_{\#2} = \frac{S_{\delta v}(f) \Big|_{\text{two devices}}}{2(V_{rms})^2} \\ = 2 S_{\delta\phi}(f) \Big|_{\text{one device}} \quad [\text{rad}^2/\text{Hz}] \quad (18)$$

The measured value is therefore divided by two in order to obtain the value for the single oscillator. A determination of the noise of each oscillator can be made if one has three oscillators that can be measured in all pair combinations.

Spectral Density of Amplitude Fluctuations

The spectral density ($S_{\delta e}(f)$) of the amplitude fluctuations of a signal follows the same general derivation previously given for the spectral density of phase fluctuations¹.

When two signals are slightly different in frequency, a slow, almost sinusoidal beat is produced at the mixer output and the peak-to-peak voltage swing is defined as (V_{ptp}).

If the two signals are now tuned to colinear phase (0 or 180° phase angle difference), the mixer output is a fluctuating voltage centered on $V_{ptp}/2$ volts. There is no requirement that the output fluctuations be small compared to $V_{ptp}/2$. In order to obtain linearity in the measurements of AM, and to make the measurement sensitive to the test oscillator only, the reference signal into the balanced mixer should be at least 10 dB greater than the test oscillator signal input to the mixers.

The normalized frequency domain measure of fractional amplitude fluctuation sidebands of a signal is called Script $\mathcal{M}(f)$. It is defined to be the ratio of the spectral density of one amplitude modulation sideband to the total signal power, at Fourier frequency difference f from the signal's average frequency ν_0 , for a single specified signal or device. The dimensionality is per hertz. Because here f is a frequency difference, the range of f is from minus ν_0 to plus infinity. Script $\mathcal{L}(f)$ and Script $\mathcal{M}(f)$ are similar functions; the former is a measure of phase modulation (PM) sidebands, the latter is a corresponding measure of amplitude modulation (AM) sidebands. We introduce the symbol Script $\mathcal{M}(f)$ in order to have useful terminology for the important concept of normalized AM sideband power.

For the types of signals under consideration, by definition the two amplitude fluctuation sidebands (lower sideband and upper sideband, at $-f$ and $+f$ from ν_0 , respectively) of a signal are coherent with each other. They are of equal intensity also. The operation of the mixer when it is driven at colinear phase is such that the amplitudes of the two AM sidebands add linearly in the output of the mixer, resulting in four times as much power in the output as would be present if only one of the AM sidebands were allowed to contribute to the output of the mixer. It has been shown that $\mathcal{M}(f)$ is related to the spectral density of the amplitude fluctuations $S_{\delta e}(f)$ by²

$$\mathcal{M}(f) = (1/2V_0^2) S_{\delta e}(f) \quad (19)$$

where V_0 is the average (nominal) amplitude also of the signal.

$$\mathcal{M}(f) = 2(\delta v \text{ (one unit)})/V_{ptp})^2 = S_{\delta v}(f)/4 V_{rms}^2 \quad (20)$$

Automated Phase Noise Measurement System

The TRW Metrology automated phase noise measurement system is shown in Figure 3. It is program controlled by the Hewlett-Packard 9830 Programmable Calculator. Each step of the calibration and measurement sequence is included in the program. The software program controls frequency selection, bandwidth settings, settling time, amplitude ranging, measurements, calculations, graphics, and data plotting. The system will be described as it is used to obtain a direct plot of Script $\mathcal{L}(f)$. The direct measurement of $\mathcal{L}(f)$ is represented by the following equation.

$\mathcal{L}(f)$ in decibels relative to 1 Hz⁻¹ =

(Noise Power Level)/(Carrier Power Level) in dB

$$-6 \text{ dB} + 2.5 \text{ dB} - 10 \log(B) - 3 \text{ dB} \quad [\text{dBc/Hz}] \quad (21)$$

The noise power is measured relative to the carrier power level and the remaining terms of the equation represent corrections that must be applied due to the type measurement and the characteristics of the measurement equipment as follows.

- o The basic measurement of noise sidebands with the signals in phase quadrature requires the -6 dB correction as explained after Equation (17).
- o The nonlinearity of the spectrum analyzer logarithmic IF amplifier results in compression of the noise peaks which, when average detected, require the +2.5 dB correction for the HP 3571A Tracking Spectrum Analyzer.
- o The bandwidth correction is required because the spectrum analyzer measurements of random noise are a function of the particular bandwidth used in the measurement.
- o The -3 dB correction is required since this is a measurement of $\mathcal{L}(f)$ using two oscillators, assuming that the oscillators are of a similar type and that the noise contribution is the same for each oscillator. If one oscillator is sufficiently superior to the other, this correction is not used.

The Calibration and Measurement Sequence

The mixer is driven with the maximum power that will result in a 50 ohm output impedance of the mixer. We used equal power inputs to the mixer when maximum noise floor was required. However, most sources were measured with a reference power of about 12 mw and a signal power level between 1 and 3 mw.

1. Measure the noise power bandwidth for each IF bandwidth setting on the Tracking Spectrum Analyzer.
2. Obtain a carrier power reference level (referenced to the output of the mixer).
3. Adjust for phase quadrature of the two signals applied to the mixer.
4. Noise power is measured at the selected Fourier frequencies, the calculations are performed, and the data is plotted (or stored) using calculator and program control (fully automated).
5. Measure and plot the system noise floor characteristics if desired.

Measurement of Noise Power Bandwidth - The precise IF noise power bandwidth of the Tracking Spectrum Analyzer must be known when performing phase noise measurements. The basic automated measurement is as follows:

- a) A tee junction is inserted at the HP 3330B synthesizer 1-MHz output and the 1-MHz signal is applied to the input of the HP 3571A Tracking Spectrum Analyzer.
- b) The calculator controls the synthesizer for the desired incremental changes in frequency. The power output is recorded for each frequency setting over the range indicated in Figure 4. The range of measurements is illustrated for equal dB values on each side of the 1-MHz center frequency. One should choose points greater than 40 dB below the carrier. We use 100 increments in frequency. Our experience has indicated that the 40 dB level and the 100

increments in frequency are not the absolute minimum permissible values.

- c) The above recorded data is plotted for each IF bandwidth as illustrated in Figure 4. The noise power bandwidth is calculated as,

$$\text{Noise bandwidth} = \frac{(P_1 + P_2 + P_3 + \dots + P_{100})\Delta f}{\text{Peak Power Reading}} \text{ [Hz]} \quad (22)$$

where Δf is the frequency increment (Hz) and the peak power is the maximum measured point obtained during the measurements. All power values are in watts.

Setting Carrier Power Reference Level - Essentially, the carrier reference power level is established at the output of the low-pass filter in Figure 3 as follows:

- The precision IF step attenuator is set to a high value to prevent overloading the spectrum analyzer (assume 50 dB as our example).
- Approximately equal power is applied to the inputs of the mixer as required to obtain the 50 ohm output impedance of the mixer in our 50 ohm system. Special techniques are required to obtain the correct impedance when lower impedance is involved.
- If the frequency of one of the oscillators can be adjusted, adjust its frequency for an IF output frequency in the range of 10 to 20 kHz.

If neither oscillator is adjustable, replace the oscillator under test with one that can be adjusted as required and that can be set to the identical power level of the oscillator under test.

- The resulting IF power level is measured by the spectrum analyzer and the measured value is corrected for the attenuator setting which was assumed to be 50 dB. The correction is necessary since this attenuator will be set to its Zero dB indication during the measurements of noise power. Assuming a spectrum analyzer reading of -40 dBm, the carrier power reference level is calculated as,

$$\begin{aligned} \text{Carrier reference power level} &= \\ 50 \text{ dB} - 40 \text{ dBm} &= 10 \text{ dBm} \end{aligned} \quad (23)$$

Phase Quadrature Adjustment - After the carrier power reference has been established, the oscillator under test and the reference oscillator are tuned to the same frequency and the original reference levels that were used during calibration are re-established. The quadrature adjustment depends upon the type of system used. Three possibilities are illustrated in Figure 3.

- If the oscillators are very stable, have high resolution tuning, and are not phase-locked, the frequency of one oscillator is adjusted for zero dc voltage output of the mixer as indicated by the sensitive oscilloscope.
- If the common reference frequency is used, as illustrated in Figure 3, then it is necessary to include a phase shifter in the line between one of the oscillators and the mixer (preferably between the attenuator and mixer). The phase shifter is adjusted to obtain zero dc output of the mixer.
- If one oscillator is phase-locked using a phase-lock loop as shown dotted in on Figure 3, the frequency of the unit under test is adjusted for zero dc output of the mixer as indicated on the oscilloscope.

The phase-lock loop is an electronic servomechanism that acts to null out any phase error between the two inputs to the mixer (phase detector). The time constant of the loop can be adjusted as needed by varying amplifier gain and RC filtering within the loop.

A loose phase-lock loop is characterized by:

- o The correction voltage varies as phase (in the short term) and phase variations are therefore observed directly.
- o The bandwidth of the servo response is small compared to the Fourier frequency to be measured.
- o The response time is very slow.

The tight phase-lock loop is characterized by:

- o The correction voltage of the servo loop varies as frequency.
- o The bandwidth of the servo response is relatively large.
- o The response time is much smaller than the smallest time interval (τ) at which measurements are performed.

NOTE: Throughout the measurement process one should check and maintain phase quadrature.

Measurements, Calculations, and Data Plots

The measurement sequence is automated except for the case where manual adjustments are required to maintain phase quadrature of the signals.

Automated measurements are executed after phase quadrature is obtained and the Precision IF Step Attenuator is set to its Zero dB indication.

As an example, our previous assumptions resulted in a carrier power level of +10 dBm. Assume that the spectrum analyzer measures -106 dBm with a 10 Hz bandwidth setting at a particular Fourier frequency. The value of spectral density which will be plotted or stored is calculated as,

$$\begin{aligned} \mathcal{L}(f) \text{ in decibels relative to } 1 \text{ Hz}^{-1} &= \\ [-106 \text{ dBm} - 10 \text{ dBm} - 6 \text{ dB} + 2.5 \text{ dB} \\ - 10 \log 10 - 3 \text{ dB}] &= -132.5 \text{ dBc/Hz} \end{aligned} \quad (24)$$

The -3 dB correction is included since it is assumed that the two units under test are similar and produce the same amount of noise.

In this example, the value of the spectral density of phase fluctuations in dB relative to one square radian per hertz is calculated as,

$$\begin{aligned} S_{\phi}(f) \text{ in dB(rad}^2/\text{Hz)} &= \\ -132.5 \text{ dBc/Hz} + 3 \text{ dB(rad}^2) &= \\ -129.5 \text{ dB(rad}^2/\text{Hz)} \end{aligned} \quad (25)$$

If one desires to plot the spectral density of frequency fluctuations, it is necessary to recall that

$$\begin{aligned} S_{\phi}(f) \text{ in dB(rad}^2/\text{Hz)} &= \\ 10 \log S_{\phi}(f) & \quad [\text{dB(rad}^2/\text{Hz)}] \end{aligned} \quad (26)$$

and that the spectral density of frequency fluctuations is

$$S_{\delta v}(f) = f^2 S_{\delta \phi}(f) \quad [\text{Hz}^2/\text{Hz}] \quad (27)$$

System Noise Floor Verification

A plot of the noise floor is obtained by repeating the automated measurements with Attenuator No. 1 set to maximum or by disconnecting the unit under test and terminating the input of the mixer with a matched load.

We use the following equation to correct for noise floor.

$\mathcal{L}(f)$ (corrected) in decibels relative to $1 \text{ Hz}^{-1} =$

$$\mathcal{L}(f) + \left[\frac{P_{\mathcal{L}(f)} - P_{\text{noise floor}}}{P_{\mathcal{L}(f)}} \right] \quad [\text{dBc/Hz}] \quad (28)$$

If adequate calculator memory is available, a set of measurements of the oscillator plus floor can be performed at selected Fourier frequencies and stored. The calculations of corrections can be made when the noise floor is measured.

The correction for noise floor can also be obtained at an earlier point in the program at Equation (13). Measurement of $S_{\delta v}(f)$ of the oscillator plus floor is obtained then $S_{\delta v}(f)$ is obtained for the noise floor only. Then,

$$\left. S_{\delta v}(f) \right|_{\text{corrected}} = \left. S_{\delta v}(f) \right|_{\text{osc. + floor}} - \left. S_{\delta v}(f) \right|_{\text{floor}} \quad [\text{V}^2/\text{Hz}] \quad (29)$$

Basics of the TRW Metrology Automated

Phase Noise Measurement Program

1. The HP 3330B Synthesizer serves as the local oscillator for the HP 3571A Tracking Spectrum Analyzer. The calculator program controls the switching of the synthesizer to the desired Fourier frequencies.
2. The Fourier frequency increments are chosen to be equal to the selected IF noise bandwidth in order to obtain a continuous spectrum plot.
3. The minimum delay time for a measurement is determined by the IF filter build-up in the spectrum analyzer. The range is from 2.5 seconds for the 3-Hz bandwidth, decreasing to 70 milliseconds for the 10-kHz bandwidth setting.
4. Video smoothing is used in order to obtain a better approximation of the mean. The program can be designed so that a large number of measurements can be taken for better estimation of the mean value. From statistical theory the confidence in an average is improved by the square root of the number of samples.
5. The IF bandwidth settings for the Fourier (offset) frequency range selections are as follows:

IF Bandwidth (Hz)	Fourier Frequency	IF Bandwidth (kHz)	Fourier Frequency (kHz)
3	10 - 400 Hz	1	40 - 100
10	400 Hz - 1 kHz	3	100 - 400
30	1 - 4 kHz	10	400 - 13 MHz
100	4 - 10 kHz		
200	10 - 40 kHz		

Program running time is 27 minutes when using 100 measurements at each Fourier measurement frequency out to

200 Hz.

6. The 60 Hz line frequency interference appears smaller than the actual amplitude if the noise corrections are applied as set forth in the noise measurement program. The corrections for the log amplifier and detection, bandwidth, and equal oscillator contribution should be removed for a plot of discrete frequencies.
7. Amplitude auto-ranging is used in the program to select the most sensitive range that does not result in overload conditions.

The low-pass filter prevents local oscillator leakage power from overloading the spectrum analyzer when baseband measurements are performed at the Fourier (offset) frequencies of interest. Leakage signals will interfere with autoranging and the dynamic range of the spectrum analyzer.

The low-noise high-gain preamplifier provides additional system sensitivity by amplifying the noise signals to be measured. Also, since spectrum analyzers usually have high values of noise figure, this amplifier is highly desirable. As an example, if the high-gain preamplifier had a noise figure of 3 dB and the spectrum analyzer had a noise figure of 18 dB, the system sensitivity at this point has been improved by 15 dB. The overall system sensitivity would not necessarily be improved 15 dB in all cases because the limiting sensitivity could have been imposed by a noisy mixer.

Figure 5 illustrates the type of plot obtained using our program and the two-oscillator measurement system. Each plot is the Normalized Phase Noise Sideband Power Spectral Density in dBc/Hz. However, any defined spectral density can be plotted as illustrated by Figure 1.

Figure 5a measurements were performed on two F.E.I. Model 1050A Disciplined Time Frequency Standards. The sample plots beneath the continuous plot are the noise floor.

Figure 5b characteristics of the Austron 1120 S Oscillator were obtained using the Austron 1125 S Multiplier to enhance the noise by 13.98 dB. Direct measurement at 5 MHz was slightly beyond the system capability. Direct measurement may be possible by incorporating a high level mixer in our system.

Figure 5c represents the characteristics of a Hewlett-Packard Model 105B Quartz Oscillator. The F.E.I. DTF Standard was used as the reference. The lowest plot is for the HP 105B and the upper plot illustrates increase in phase noise when the signal was amplified using a Hewlett-Packard 5087A Distribution Amplifier.

Figure 5d is the phase noise plot obtained using two identical 250 MHz Surface Acoustic Wave (SAW) Oscillators developed at TRW. The roll-off in the range from 10 Hz out to about 200 Hz was due to inadequate isolation between the two oscillators. The slight injection locking was also noted when fine tuning one oscillator. This illustrates a problem of measuring in close to the carrier since the effective locking loop bandwidth will depend upon the isolation of the oscillators.

Figure 5e shows measurements performed on two HP 8660C Synthesizers, at 5 MHz. The reference oscillator of one was locked to the reference oscillator in the other unit. Therefore, the phase noise plot does not slope upward from about 500 Hz to 10 Hz on the plot. The slope upward was verified using the single-oscillator measurement technique which will be discussed later. The phase noise reduction by the phase-lock loop of the synthesizer is shown and one notes where

the loop bandwidth ends and where the characteristics of the source oscillator are being measured.

Figure 5f shows measurements performed on a Fluke 6160B Synthesizer with the F.E.I. 7507A-1 Frequency Standard as the reference. The measurements were performed at 100 MHz. The characteristics discussed previously are now shown. The top plot at the lower Fourier frequencies was obtained with the oscillator "free-running". The lower plot was performed when the reference oscillator of the synthesizer was phase-locked to the F.E.I. Frequency Standard Reference Oscillator.

Phase Noise Measurements Using Delay Line FM Discriminators

Frequency fluctuations are measured directly using FM discriminator techniques (References 5, 6, and 7). One of the important advantages of this type system is that the phase noise characteristics of a single oscillator can be measured without the requirement of a similar or better source as a reference.

The delay line yields a phase shift by the time the signal arrives at the balanced mixer. The phase shift depends upon the instantaneous frequency of the signal. The presence of frequency modulation (FM) on the signal gives rise to differential phase modulation (PM) at the output of the differential delay and its associated (non-delay) reference line. This relationship is linear if the delay (τ_d) is non-dispersive. This is the property which allows the delay line to be used as an FM discriminator. In general, the conversion factors are a function of the delay (τ_d) and the Fourier frequency (f) but not the carrier frequency. The delay of the transmission line and variable length airline as a function of phase noise measurement at various offset frequencies will be discussed later. However, the maximum sensitivity of the transmission line discriminator depends upon the attenuation value of the delay line at the carrier frequency.^{5,7} For maximum sensitivity, the optimum length of the delay line for the system shown in Figure 6 occurs when the total attenuation is approximately one neper (one neper = 8.686 dB). It represents the one-way loss of the delay line in Figure 6a and the round-trip loss of the delay line shown in Figure 6b. The optimum delay line length is determined at a particular selectable carrier frequency. However, since the attenuation varies slowly (approximately proportional to the square root of frequency), this characteristic allows near optimum operation over a considerable frequency range without appreciable degradation in the measurements.

General Theory

The measurement system is shown in Figure 6. Two delay line configurations are illustrated. The signals experience the one-way delay of the delay line in Figure 6a.^{5, 6} The FM discriminator of Figure 6b is a reflective type used by Ashley et al⁷ and it is noted that the signals experience the round-trip and resultant two-way loss in the delay line.

With adequate drive power we have not limited the delay line in this system to one neper. Delay lines with attenuation greater than 11 dB were used in some of our measurements. The resulting noise floor was more than 10 dB below the measured values of phase noise.

The calculator program performs the same functions as outlined for the measurement system of Figure 3.

The system calibration and measurement procedure is outlined as follows:

o Measure IF Noise Power Bandwidths

The IF noise power bandwidth is measured for each IF bandwidth setting of the spectrum analyzer as set forth in a previous section of this paper.

o Establish System Power Levels

The oscillator under test is connected and Attenuators 1, 2, and 3 are used to set the equal power levels at the mixer inputs. The inputs were from +7 to +10 dBm for the particular mixer used in our system.

o Quadrature Adjustment

The phase shifter is adjusted for Zero volts DC at the output of the mixer as indicated on the oscilloscope connected as shown. This establishes the quadrature condition for the two inputs to the mixer. This quadrature condition is continuously monitored and is adjusted if necessary.

o Discriminator Calibration

The calibration factor of the discriminator should be established at selected Fourier frequencies over the range of measurements. If the calibration factor is not constant, appropriate measured values should be used in the automated measurement program. The measurement and calibration will be described in terms of a 20 kHz modulation frequency.

The calibration factor is defined as:

$$CF = \Delta v_{rms} / V_{rms} \quad [Hz/V] \quad (30)$$

Where Δv_{rms} is the rms frequency deviation of the carrier due to the intentional modulation and V_{rms} is the spectrum analyzer voltage measurement of the modulation sideband.

$$\Delta v_{rms} = \Delta v_{peak} / \sqrt{2} = m (f_m) / \sqrt{2} \quad [Hz] \quad (31)$$

The calibration factor of the discriminator is calculated as:

$$CF = m (f_m) / \sqrt{2} V_{rms} \quad [Hz/V] \quad (32)$$

The calibration factor of this system has been found to be constant to about 400 kHz which was the limit due to resolution in setting the carrier null.

The calibration factor is established as follows:

- a) Replace the oscillator under test with a signal generator or oscillator that can be frequency modulated.
The power output and operating frequency of the generator must be set to the same precise frequency and amplitude values as the oscillator under test.
- b) Select a modulation frequency of 20 kHz and increase the modulation until the carrier is reduced to the first Bessel null of the carrier as indicated on the spectrum analyzer connected to Coupler No. 1. One can choose any other convenient setting that will produce a known modulation index (m).
- c) Tune the HP 3571A Tracking Spectrum Analyzer to the modulation sideband frequency (20 kHz). Maintain the phase quadrature of the two signals applied to the balanced mixer.
- d) The Tracking Spectrum Analyzer is now displaying a power reading that corresponds to the

- d) modulation frequency. This reading must be corrected for the dB setting of Attenuator No. 4 in order to calculate the calibration factor of the discriminator. Attenuator No. 4 will be set to Zero dB reference indication during the measurements of Fourier frequency noise power. This results in an increase of sideband power at the mixer which must be accounted for in the calculation of the calibration factor.

As an example, if the attenuator was set to 50 dB as proposed, and the spectrum analyzer reading of baseband power is -35 dBm, the corrected power is then

$$P \text{ (dBm)} = 50 \text{ dB} - 35 \text{ dBm} = +15 \text{ dBm} \quad (33)$$

- e) The discriminator calibration factor can now be calculated, since this power in dBm can be converted to the corresponding rms voltage using the following equation.

$$V_{\text{rms}} = \sqrt{10^{P(\text{dBm})/10} / 1000 \times R} \quad [\text{V}] \quad (34)$$

where R is 50 ohms in this system.

- f) The discriminator calibration factor is calculated as,

$$\begin{aligned} \text{CF} &= m \cdot f_m / \sqrt{2} V_{\text{rms}} \\ &= 2.405 \cdot f_m / \sqrt{2} V_{\text{rms}} \quad [\text{Hz/V}] \quad (35) \end{aligned}$$

since 2.405 is the modulation index (m) for the first Bessel carrier null as used in this technique. The modulation frequency is f_m .

Measurement and Data Plotting

- a) Connect the unit under test in place of the modulated signal generator and readjust quadrature at the signals if necessary.
- b) Set Attenuator No. 4 to its Zero dB indication.

The measurements, calculations, and data plotting are completely automated. The calculator program selects the Fourier frequency, auto-ranging is performed, bandwidth is set, and measurements of Fourier frequency power are performed by the tracking spectrum analyzer.

Each Fourier frequency noise power reading P_n (dBm) is converted to the corresponding rms voltage designated as $\delta v_{1\text{rms}}$.

$$\delta v_{1\text{rms}} = \sqrt{10^{(P_n(\text{dBm}) + 2.5)/10} / 1000} \times R \quad [\text{V}] \quad (36)$$

the rms frequency fluctuations are calculated as,

$$\delta v_{\text{rms}} = \delta v_{1\text{rms}} \times \text{CF} \quad [\text{Hz}] \quad (37)$$

The spectral density of frequency fluctuations is calculated as,

$$S_{\delta v}(f) = (\delta v_{\text{rms}})^2 / B \quad [\text{Hz}^2/\text{Hz}] \quad (38)$$

where B is the measured IF noise power bandwidth of the spectrum analyzer.

The spectral density of phase fluctuations is calculated as,

$$S_{\delta \phi}(f) = S_{\delta v}(f) / f^2 \quad [\text{rad}^2/\text{Hz}] \quad (39)$$

The normalized phase noise sideband power spectral density is calculated as,

$$\mathcal{L}(f) = 1/2 \text{ rad}^2 S_{\delta \phi}(f) \quad [\text{Hz}^{-1}] \quad (40)$$

Script $\mathcal{L}(f)$ expressed in decibels relative to 1 Hz^{-1} is plotted in real time in our program. However, the data can be stored and the desired spectral density can be plotted in other forms.

Noise Floor Measurements

The system noise floor can be plotted by setting Attenuator No. 3 to maximum and repeating the automated measurements. We increase the LO power to obtain essentially the same mixer output impedance which exists when both signals are applied. When using the 12 mw and 3 mw levels, there is no difference in the noise floor with and without the 3 mw signal. However, in systems which require different impedance levels, it is necessary that the total input power (Signal and LO) be adjusted to produce the required mixer output impedance. This will require a method of measuring the mixer output impedance and re-establishing the particular value when measuring the noise floor. In the calculation process in the program, the rms voltage corresponding to the noise floor is designated $\delta v_{2\text{rms}}$.

In our program the noise floor is checked only at selected Fourier frequencies over the range of measurement.

A correction for the noise floor requires a measurement of the rms voltage of the oscillator and floor $\delta v_{1\text{rms}}$ and a measurement of the noise floor rms voltage $\delta v_{2\text{rms}}$. These voltages are used in the following equation to obtain the corrected value.

$$\delta v_{\text{rms}} = \sqrt{(\delta v_{1\text{rms}})^2 - (\delta v_{2\text{rms}})^2} \quad [\text{V}] \quad (41)$$

The corrected value δv_{rms} is then used in the calculation of frequency fluctuations. If adequate memory is available, each value of $\delta v_{1\text{rms}}$ can be stored and used after the other set of measurements is performed at the same Fourier frequencies.

Figure 6b illustrates the FM discriminator used by Ashley et al⁷. The circulator, double stub tuner, delay line and adjustable length air-line form the FM discriminator. The double stub tuner and adjustable line are used to null out the carrier.

The carrier nulling procedure makes the discriminator relatively immune to incidental AM on the signal being tested. The system is not restricted by burnout characteristics of the crystals in the mixer.

The quadrature adjustment differs from the system of Figure 6a which uses the one-way length of the delay line. During the calibration process the phase shifter is adjusted for maximum output indication on the Tracking Spectrum Analyzer at the 20 kHz modulation frequency. This sets the phase quadrature of the two signals applied to the balanced mixer.

The phase noise plots of Figure 7 were obtained with the TRW Metrology program and the measurement system shown in Figure 6a. The phase noise

plot follows the power law process to about 700 kHz. Beyond this point the delay line discriminator is not calibrated. A plot of $S_{\delta v}(f)$ has nulls at every multiple of f_d , where f_d is the Fourier frequency corresponding to $1/t_d$, and t_d is the total differential delay of the FM discriminator.⁵

Figure 7a measurements were performed using two different delay lines. The delay line used in measurements of the 600 MHz fundamental frequency was about 500 ns as indicated by the first null at about 2 MHz. The delay line used in the measurements of the 2.4 GHz (multiplied output) had about 250 ns delay as indicated by the first null at about 4 MHz. Note that the theoretical enhancement and measured data agree within the resolution of the plot.

APPENDIX A

Basic Modulation Theory and Spectral Density

Applying a sinusoidal frequency modulation (f_m) to a sinusoidal carrier frequency (v_o) produces a wave that is sinusoidally advanced and retarded in phase as a function of time. The instantaneous voltage is expressed as,

$$V(t) = V_o \sin [2\pi v_o t + \Delta\phi \sin 2\pi f_m t] \quad (1)$$

where $\Delta\phi$ is the peak phase deviation caused by the modulation signal.

The first term inside the brackets represents the linearly progressing phase of the carrier. The second term is the phase variation (advancing and retarded) from the linearly progressing wave.

The effects of modulation can be expressed as residual FM noise in which modulation by a single sinusoidal signal produces a peak frequency deviation of the carrier (v_o).

$$\Delta v_o = \Delta\phi \cdot f_m \quad (2)$$

$$\Delta\phi = \Delta v_o / f_m \quad (\text{peak}) \quad (3)$$

This ratio of peak frequency deviation to modulation frequency is called modulation index (m) so that $\Delta\phi = m$ and

$$m = \Delta v_o / f_m \quad (4)$$

The frequency spectrum of the modulated carrier contains frequency components (side-bands) other than the carrier. For small values of modulation index ($m < 1$), as is the case with random phase noise, only the carrier and first upper and lower sidebands are significantly high in energy.

The effects of modulation can also be expressed as single sideband phase noise in which the peak phase deviation is $\Delta\phi = m$. The ratio of the amplitude of either sideband (single sideband) to the amplitude of the carrier, V_o is:

$$V_{sb}/V_o = m/2 \quad (5)$$

This ratio is expressed in dB and referred to as dB below the carrier for the given bandwidth.

$$\begin{aligned} \left. \frac{V_{sb}}{V_o} \right|_{dB} &= 20 \log(m/2) = 20 \log(\Delta v_o / 2f_m) \\ &= 10 \log \left[\frac{m}{2} \right]^2 = 10 \log(\Delta v_o / 2f_m)^2 \quad [\text{dBc/B}] \quad (6) \end{aligned}$$

If the frequency deviation is given in terms of its rms value then,

$$\Delta v_{rms} = \Delta v_o / \sqrt{2} \quad (7)$$

Equation 6 now becomes

$$\begin{aligned} \left. \frac{V_{sb}}{V_o} \right|_{dB} &= \mathcal{L}(f) = 20 \log(\Delta v_{rms} / \sqrt{2} f_m) \\ &= 10 \log(\Delta v_{rms} / \sqrt{2} f_m) \quad (8) \end{aligned}$$

The ratio of single sideband to carrier power in dBc/Hz is,

$$\mathcal{L}(f) = 20 \log \left[\frac{\Delta v_{rms}}{f_m} \right] - 3 \text{ dB} \quad [\text{dBc/Hz}] \quad (9)$$

and

$$S_{\delta\phi}(f) = 10 \log \left[\frac{\Delta v_{rms}}{f_m} \right] \quad [\text{dB}(\text{rad}^2/\text{Hz})] \quad (10)$$

i.e., in decibels relative to $1 \text{ rad}^2/\text{Hz}$.

The interrelationships of modulation index, peak frequency deviation, rms frequency deviation and spectral density of phase fluctuations can be found from

$$\begin{aligned} \left[\frac{m}{2} \right]^2 &= \left[\frac{\Delta v_o}{2 f_m} \right]^2 = \left[\frac{\Delta v_{rms}}{\sqrt{2} f_m} \right]^2 \\ \text{or} \quad &= 10 \mathcal{L}(f)/10 = S_{\delta\phi}(f)/2 \quad (11) \end{aligned}$$

$$\frac{m}{2} = \frac{\Delta v_{rms}}{\sqrt{2} f_m} = \sqrt{10 \mathcal{L}(f)/10} = \sqrt{S_{\delta\phi}(f)/2} \quad (12)$$

and

$$\begin{aligned} m &= \frac{\Delta v_o}{f_m} = \frac{2 \Delta v_{rms}}{\sqrt{2} f_m} = 2 \sqrt{10 \mathcal{L}(f)/10} \\ &= 2 \sqrt{S_{\delta\phi}(f)/2} \quad (13) \end{aligned}$$

The basic relationships are plotted in Figure 1.

Discussion

Identical phase noise plots were obtained using the two discriminator systems. Measurements performed with the single-oscillator technique and the two-oscillator technique agree over the range out to about 40 percent of the Fourier frequency of the first null of the plot obtained with the single-oscillator (FM discriminator) system. The FM discriminator is uncalibrated beyond the described 40 percent area as indicated in Figure 7.

When using the two-oscillator technique accurate measurements close to the carrier can be limited depending upon the isolation between the two sources. This was clearly indicated in the data plots of Figure 5.

In order to obtain the lowest noise floor, the mixers used in these systems were driven with equal signals to the RF and LO ports. The nominal IF output impedance of the mixers was 50 ohms and we did not experience system sensitivity problems with a broad range of total input power to the mixer. Mismatch problems were not critical when the mixers were driven for IF output impedance values in the range of 30 to 70 ohms.

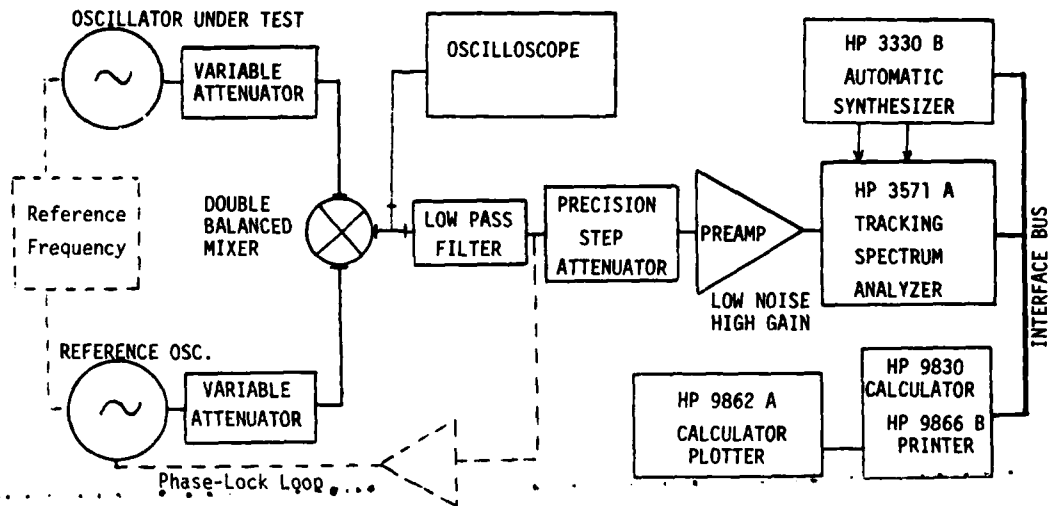


FIGURE 3. PHASE NOISE MEASUREMENT SYSTEM USING THE TWO - OSCILLATOR TECHNIQUE.

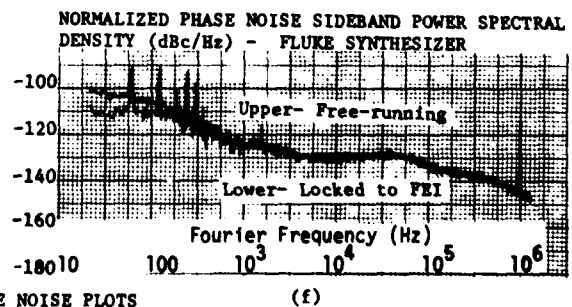
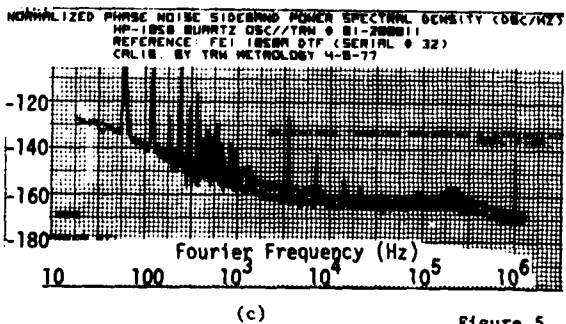
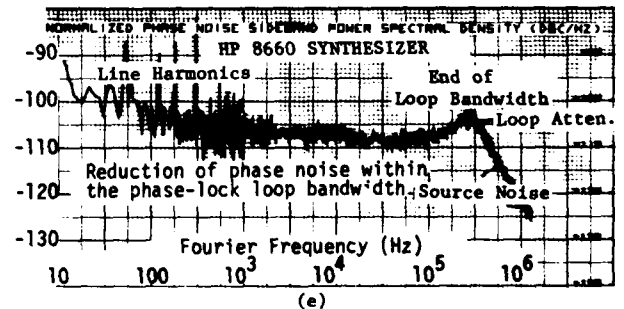
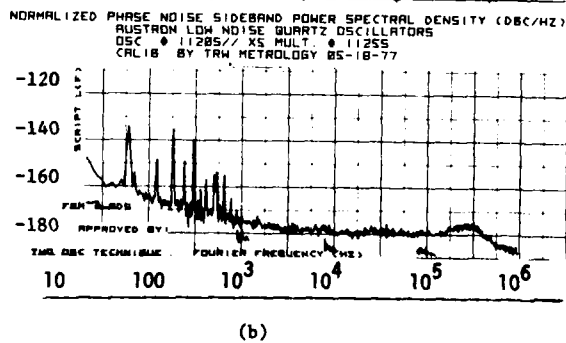
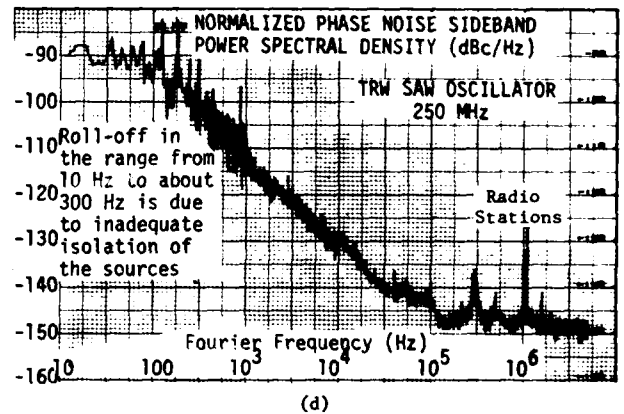
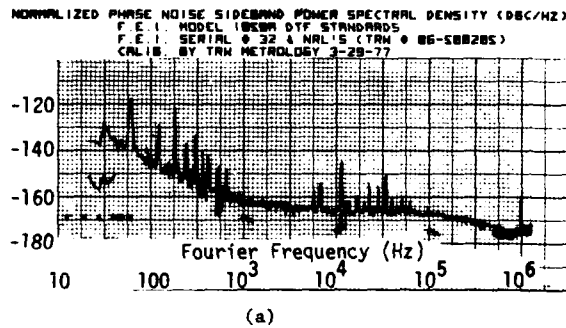


Figure 5. PHASE NOISE PLOTS

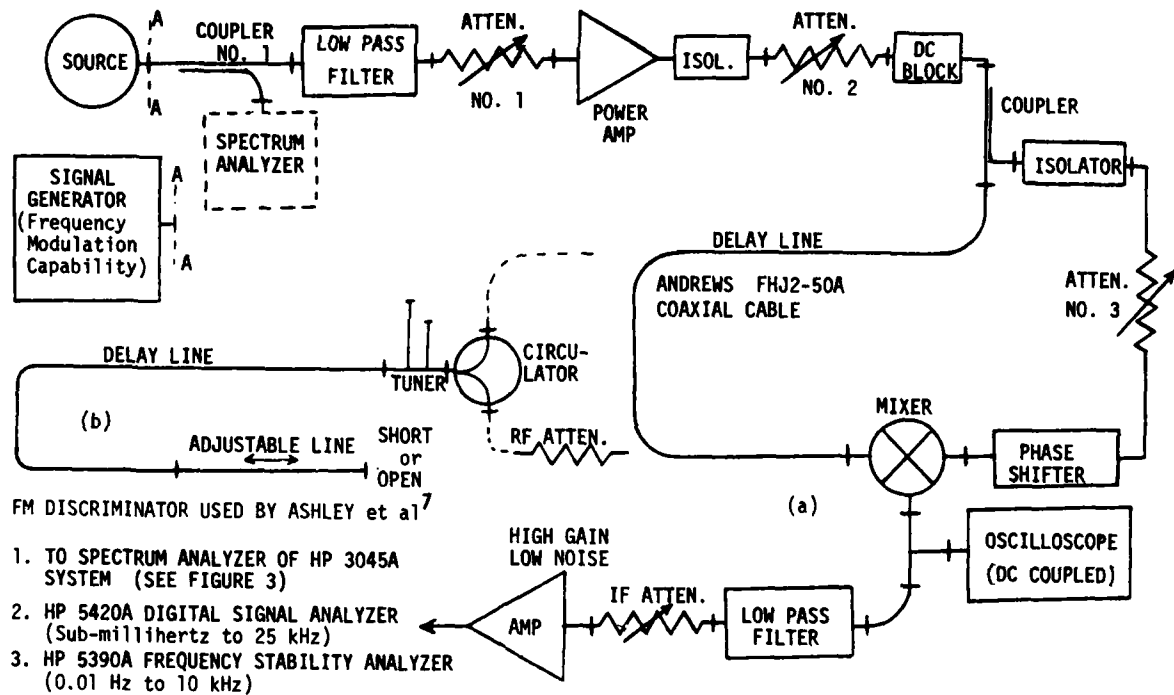


FIGURE 6. SINGLE - OSCILLATOR PHASE NOISE MEASUREMENT SYSTEM USING THE DELAY LINE AS AN FM DISCRIMINATOR.

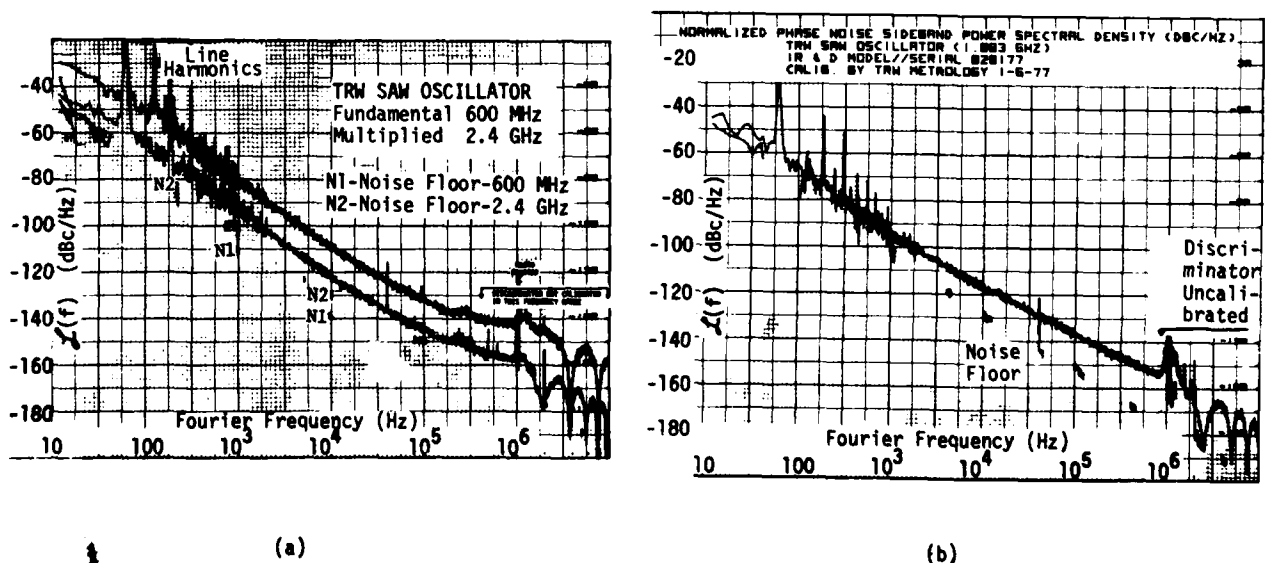


FIGURE 7. NORMALIZED PHASE NOISE SIDEBAND POWER SPECTRAL DENSITY OF TRW DEVELOPED SURFACE ACOUSTIC WAVE OSCILLATORS MEASURED USING THE SINGLE - OSCILLATOR TECHNIQUE.

CURRENT DEVELOPMENTS IN SAW OSCILLATOR STABILITY*

T. E. Parker
Raytheon Research Division
Waltham, Massachusetts

Abstract

The surface acoustic wave (SAW) controlled oscillator has been the subject of considerable research over the past few years and a significant improvement in the understanding and performance of SAW oscillators has been achieved. The frequency stability of an oscillator is a key parameter and will be discussed in three regimes: (1) short-term stability - noise, (2) medium-term stability - temperature effects, and (3) long-term stability - aging.

The equations for calculating the FM noise for both resonator and delay line oscillators will be reviewed. For both the resonator and delay line, the material propagation loss ultimately limits the noise characteristics of the SAW device. Theoretical limits are discussed and it is shown that an ideal delay line oscillator (SAW insertion loss is due only to propagation loss) has a lower close in noise level than an ideal resonator by about 3 dB. However, an ideal delay line is extremely difficult to implement, and in practice a resonator oscillator will almost always give a better noise performance. Results of FM noise measurements on resonators and delay line type oscillators are presented. Though resonator oscillators have better noise performance, the delay line oscillator is sometimes attractive because of its larger tuning range relative to its delay time.

Although research on new SAW materials is continuing, no significant advances in SAW oscillator temperature stability have been made. A quick review of the temperature dependence of ST-cut quartz and the $\text{SiO}_2/\text{LiTaO}_3$ structure is presented.

Aging has been a major problem with SAW controlled oscillators, with aging rates on the order of -1 to -10 ppm/month having been reported. These rates were observed, however, on devices which were packaged in a relatively crude manner using RTV mounting. Recent results using no organics and high-vacuum sealing have shown positive aging rates as low as +0.25 ppm/month.

Introduction

Over the last few years there has been considerable improvement in SAW oscillator stability and increased understanding of the phenomena which affect frequency stability. The use of SAW resonators has given improved short-term stability (noise) and equations will be presented for calculating FM noise for both resonators and delay lines. From these equations the best design for optimum noise performances can be determined. Phase noise measurements for both delay lines and resonators will be presented.

Improvements in long-term stability (aging) have also occurred, largely due to better packaging techniques. Cleaning and sealing techniques similar to those used with bulk wave resonators have produced

oscillators with aging rates as low as +0.25 ppm/month. Aging data for a number of oscillators is presented.

Medium-term stability (temperature dependence) has not seen significant improvement over previously reported results; therefore only a brief review is presented.

In the following discussion all results were obtained with two port delay lines and resonators used in a simple feedback oscillator as shown in Fig. 1.

Short-Term Stability - FM Noise

The feedback oscillator in Fig. 1 inherently operates with the amplifier in a saturated condition and therefore the AM noise is suppressed. Consequently, the dominant noise is FM. A convenient way of expressing FM noise is the single-sideband noise power spectrum. Equation 1 is based on an expression by Leeson¹ and gives the FM single-sideband noise power relative to the carrier in a 1 Hz bandwidth as a function of the offset or modulation frequency, $\Delta\omega$.

$$\left(\frac{P_{SB}}{P_c} \right)_{dBc} = 10 \log \left\{ \left(\frac{N^2 G F K T}{P_c} \right) \left(\frac{\omega_g}{(\Delta\omega)^3 \tau^2} + \frac{1}{(\Delta\omega)^2 \tau^2} + 1 \right) \right\} \quad (1)$$

- where
- N = multiplication factor
 - G = amplifier gain \approx SAW insertion loss
 - KT = thermal energy
 - F = amplifier noise figure
 - P_c = oscillator power at amplifier output
 - τ = phase slope (group delay) of SAW device
 - ω_g = flicker frequency
 - $\Delta\omega$ = offset or modulation frequency

Equation 1 is not an exact expression, but gives only the basic characteristics of the noise power spectrum. There is some disagreement in the literature^{1,2,3} over a factor of two which may appear in the term in the left hand parenthesis but Eq. 1 has been found to be accurate to within a few dB in most cases. The right-hand term of the three terms enclosed in the right-hand parenthesis represents the thermal floor which is independent of modulation frequency. As one would expect, this level depends on the amplifier gain and noise figure as well as KT and the carrier power P_c . If multiplication is used, this also affects the thermal floor by making N larger than one. The middle-term includes the effect of the group delay or phase slope in the noise power spectrum, and results from a frequency modulation of the oscillator caused by phase noise at the amplifier input. Below some frequency,

* Work co-sponsored by National Aeronautics and Space Administration, Goddard Space Flight Center, Contract No. NAS5-23701.

ω_a , the phase noise begins to increase as $1/\Delta\omega$ and this is the flicker noise contribution represented by the left-hand term. This results in a further increase in the spectrum slope to 30 dB/decade. It has been assumed in Eq. 1 that $\omega_a < 1/\tau$.

From Eq. 1 it is evident that several steps can be taken to minimize the noise in an oscillator. Specifically, G , F , and N should be minimized and can be controlled to some extent by SAW design and the choice of amplifiers. Also, P_c should be made as large as possible. ω_a is not a design parameter but generally is affected by fabrication techniques in both the SAW device and amplifiers. It is desirable that ω_a be as small as possible, but the value of ω_a is not a design variable and is generally not predictable. To further reduce the close-in noise, τ should be maximized, but τ cannot be increased without limit since this will increase the SAW insertion loss and therefore require a larger G . The value of τ which gives the lowest close-in noise level can be determined from Eq. 1, but the value is different for delay lines and resonators.

For a delay line the insertion loss (in dB) increases linearly with τ and it is easy to show that the close-in noise level, N_{FM} , is proportional to

$$N_{FM} \propto \frac{e \left(\frac{\ln 10}{10} \alpha \tau \right)}{\tau^2}$$

where α is the surface wave propagation loss in dB/ μ sec and τ is the group delay in μ sec. Differentiating this relation and setting it to zero gives the result that the optimum value of τ is

$$\tau_{opt} = \frac{8.69}{\alpha}$$

which gives an insertion loss (due only to propagation loss) of 8.69 dB.

For a resonator the insertion loss is a function of the loading, or matching, and is given by

$$I.L. (dB) = 10 \log \left(\frac{1}{1 - \frac{Q_L}{Q_U}} \right)^2$$

where Q_L and Q_U are the loaded and unloaded Q 's. Proceeding in the same manner as with the delay line it can be shown that for a resonator

$$\tau_{opt} = \frac{Q_U}{\omega}$$

which is equivalent to

$$Q_L = \frac{Q_U}{2}$$

since for a resonator

$$\tau = \frac{2Q_L}{\omega}$$

where ω is the oscillator frequency. The insertion loss of a resonator tuned for optimum delay is 6 dB.

Acoustic propagation loss directly affects the

optimum delay of delay lines, as shown above, and it has a similar effect on resonators because it limits the unloaded Q , Q_U , which is equivalent to the material Q of reference 4. Acoustic propagation loss increases⁵ as ω^2 and the resulting optimum delay (or phase slope) as a function of oscillator frequency is shown in Fig. 2 for delay lines and resonators on ST cut quartz. Obtaining the optimum delay with a delay line requires only that the device have the proper length, though this can be a problem at low frequencies where the length gets very large. For a resonator, losses in the grating must be minimized so that the unloaded Q is close to the material limit. At this point the device is then matched to give a loaded Q equal to $Q_U/2$ and an insertion loss of 6 dB.

It is interesting to note that for an ideal delay line (that is, a delay line in which there are no transducer losses and the net insertion loss is due only to propagation loss), the close-in noise level is about 3 dB lower than that of a perfect resonator. However, an ideal delay line is very difficult to implement and generally a SAW resonator will give better noise performance than a typical delay line.

One advantage that the delay line does have, however, is that it is inherently more wideband than the resonator. The 3 dB bandwidth of a single-mode delay line is about three times larger than a resonator with the same group delay. Also the group delay of the delay line is essentially constant over this bandwidth, whereas the resonator group delay decreases as it is tuned off resonance. The delay line is also easily adapted to multi-mode operation and oscillators with a percent or more tunability have been demonstrated which have very respectable noise characteristics.^{6,7}

Generally, it can be concluded that for very narrow-band, or fixed-frequency applications, the resonator-type oscillator will give the best noise performance. For applications where tunability and linearity are important, the delay-line-type oscillator may give the best performance.

To illustrate typical noise characteristics of SAW oscillators, measurements have been made on a 401-MHz delay line and a 310-MHz resonator. Figure 3 shows measured and calculated single-sideband FM noise for these two oscillators and also lists the values of the various parameters. The flicker frequency, $f_a = \omega_a / 2\pi$, was determined from the experimental data and used in the calculated curves. In both cases the overall agreement is good. The dashed lines were inserted in the resonator data to emphasize the 20 and 30 dB/decade slopes and to more clearly show how the flicker frequency is determined by the intersection of the two slopes. The bump in the resonator noise data at 300 KHz is caused by secondary resonances in the SAW device which are about 3 dB higher in insertion loss than the main resonance. The secondary resonances were intentionally designed into the SAW device and can be removed with a new design. The noise level for the delay line oscillator drops below theory for $\Delta f > 100$ KHz because the theory does not include the effect of destructive interference which causes the noise level to be suppressed³ somewhat at offset frequencies in the vicinity of $1/2\tau$. This does not occur in the resonator because the amplitude response of a resonator falls off more rapidly than for a delay line.

Another way of characterizing the frequency stability of an oscillator is to plot the fractional frequency stability as a function of measurement period, T . Figure 4 shows results for the 401-MHz delay line as measured with a Hewlett-Packard 5360A computing counter. The solid line was calculated⁸ from the noise spectrum shown in Fig. 3. The increase in measured

frequency deviation for $T = 1$ and 10 sec is probably due to residual frequency drift due to temperature changes.

Returning to Eq. 1, another factor which affects short-term stability is the amount of multiplication which is used. Using an oscillator with a high fundamental frequency can reduce the value of N , but, as shown in Fig. 2, this also reduces the optimum value of τ . Since τ decreases as ω^2 , it is obvious that a high oscillator frequency does not result in improved close-in noise. This is shown in Fig. 5, which gives the calculated single-sideband FM noise at 100 Hz for a 10 GHz system as a function of the oscillator frequency. Results for both an ideal delay line and a resonator are shown. In both cases flicker noise has not been included. The figure clearly shows that the lowest close-in noise is obtained if a low frequency oscillator is used with the necessary multiplication needed to reach 10 GHz. However, two points should be noted. First, the opposite relation holds for far-out noise. The thermal floor is essentially independent of oscillator frequency, but does increase with the multiplication factor. Thus, good close-in performance is obtained at the expense of increased far-out noise. Second, there may be many practical reasons for not wanting to use multipliers. These may include reliability, power consumption, and spurious suppression. To illustrate the difference in noise characteristics, Fig. 6 shows the single-sideband FM noise of a commercial 5 MHz bulk wave resonator⁹ multiplied to 310 MHz along with a 310-MHz surface acoustic wave resonator (same device as shown in Fig. 3). Above 1 KHz the surface wave oscillator is better but below 1 KHz the bulk wave oscillator is better. The group delay of the surface wave oscillator is only about half of the theoretical limit, so there is room for improvement here and also the flicker frequency may be reducible with amplifiers or SAW fabrication. The dashed line in Fig. 6 is an extension of the 20 dB/decade region to illustrate the SAW oscillator performance without flicker noise. Nevertheless, even with significant improvement, it is doubtful the SAW oscillator could be made to have lower close-in noise than the multiplied bulk wave oscillator.

Medium-Term Stability - Temperature Effects

Although there is considerable research being done on new materials,¹⁰ there have been no significant improvements in oscillator temperature stability. The only two demonstrated materials for temperature stable SAW oscillators are ST-cut quartz and the $\text{SiO}_2/\text{LiTaO}_3$ overlay structure.¹¹ Figure 7 shows the temperature dependence of oscillators made with these two materials. ST-cut quartz is the most commonly used material since it is the simplest to use, but the $\text{SiO}_2/\text{LiTaO}_3$ structure offers better performance at the expense of more complicated fabrication techniques. In addition to new materials, electronic techniques can be used to stabilize an oscillator but very little has been done at this time.

Long-Term Stability - Aging

Aging tests have been going on at Raytheon for the past two-and-a-half years and it has become obvious that the observed aging rates are largely related to cleaning and packaging. Our very first oscillator was mounted in an unsealed package, and it quickly became evident that the frequency went up and down with the humidity. Subsequent devices were mounted in RTV and solder-sealed in gold-plated flatpacks.¹² These devices were baked at 80° C for 6 to 8 hours in a dry nitrogen atmosphere before final sealing. Figure 8 shows the aging data for 13 devices. Numbers 1 through 8 were mounted as just described. All of

these devices operated near 300 MHz but devices 2, 4, and 5 were made with the $\text{SiO}_2/\text{LiTaO}_3$ structure. The other five were made on ST-cut quartz. Devices 6 and 7 had chrome-gold transducers while all the others had aluminum metallization. Device number 8 was the only resonator. Of the eight devices packaged with RTV, all aged downward in frequency. An accumulation of foreign material would cause the frequency to decrease so it was postulated that organic components from the RTV were condensing on the crystal surface.

To eliminate the use of RTV, it was decided to use gold wires to strap the crystal in place. By notching the sides of the crystal and passing the straps through the notches, the crystal could be held in place without subjecting it to large amounts of stress. Though this mounting scheme is not as immune to vibration as that using RTV, it involves no organics and can be made very clean. The technology of bulk wave resonator fabrication was applied wherever possible. After mounting in gold-plated flatpacks, the devices were then placed in a high-vacuum chamber (2×10^{-7} torr) and were baked at 200° C for several hours. The final sealing was done by heating the package to 310° C to melt a gold-tin preform. Devices 9, 10, 12, and 13 were sealed in this manner and their aging characteristics are also shown in Fig. 8. These devices are all delay lines, with numbers 9 and 10 operating at 310 MHz and 12 and 13 operating at 401 MHz. All of the devices packaged under high vacuum and with no organics have aged upwards. Generally, the four devices packaged in evacuated flatpacks have shown aging characteristics that are nearly logarithmic with time, whereas the RTV-mounted devices were more nearly linear. Consequently, the evacuated devices show a more rapidly decreasing aging rate.

Device number 11 was also packaged under high vacuum, but it is unique because it was packaged in a standard HC-36/U cold weld enclosure. The packaging was accomplished at the United States Army Electronics Command in Ft. Monmouth, New Jersey, through the courtesy of Ted Lukaszek and Bill LeBus. Device 11, like 9 and 10, had aluminum transducers underlaid with a flash of chrome. Number 11 was also strapped onto a stainless steel backing plate with gold wires as shown in Fig. 9. The final cleaning was accomplished at ECOM using standard chemical and UV procedures.¹³ The device was placed in the high vacuum system ($\sim 10^{-8}$ torr) and baked at 150° C for one hour. The package was then sealed by cold weld. The packaging facility at ECOM is used for sealing state-of-the-art bulk wave resonators, so it is reasonable to assume that device 11 is packaged under conditions similar to that of good bulk wave devices. Device 11 is aging upward like the other devices packaged under high vacuum, but the total frequency drift is significantly less. Unlike the other evacuated devices, number 11 is not aging logarithmically but it is only somewhat sublinear.

In addition to the various packaging schemes that were used, Fig. 8 shows that several other parameters have been varied. In particular the two chrome-gold devices and the resonator show the lowest aging of the RTV mounted devices. This is an interesting observation but it is difficult to draw any conclusions since the packaging was not clean. Other parameters which may affect aging are the type of surface polish and the quality of the quartz material. Devices 12 and 13 were made on substrates having a chemical-mechanical polish, and device 13 was made on a "pure Z growth" material rather than the standard ultrasonic grade used for all the others. Future experiments are planned to evaluate such parameters as types of metallization, surface polish and quality of quartz, but it is difficult or impossible to reach

definite conclusions about these parameters until packaging and cleaning variables are eliminated. Experiments are in progress now to evaluate the aging of the $\text{SiO}_2 / \text{LiTaO}_3$ structure under clean packaging conditions.

Acknowledgements

The author would like to thank Ted Lukaszek and Bill LeBus of the United States Army Electronics Command for providing the packaging of one of the devices. Thanks are also given to C. Dunnrowicz and A. Stankiewicz for packaging and testing of the other devices.

References

1. D.B. Leeson, "Short Term Stable Microwave Sources," *Microwave Journal*, p. 59-69 (June 1970).
2. D.B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum," *Proc. IEEE* 54, 329 (February 1966).
3. M. Lewis, "The Surface Acoustic Wave Oscillator - A Natural and Timely Development of the Quartz Crystal Oscillator," *Proc. 28th Annual Frequency Control Symp.*, 304 (May 1974).
4. D.T. Bell and R.C.M. Li, "Surface Acoustic Wave Resonators," *Proc. IEEE* 64, 711 (May 1976).
5. A.J. Slobodnik, "A Review of the Properties of Acoustic Surface Wave Materials," *Proc. Symp. on Optical and Acoustical Microelectronics*, 23, 205 (April 1974).
6. T.E. Parker, "A SAW Oscillator using Two Acoustic Paths," *Proc. of 1976 IEEE Ultrasonics Symposium*, 243 (Sept. 1976).
7. I. Browning, J. Crabb and M.F. Lewis, "A SAW Frequency Synthesizer," *Proc. of 1975 IEEE Ultrasonics Symposium*, 245 (Sept. 1975).
8. J. Rutman, "Relations Between Spectral Purity and Frequency Stability," *Proc. 28th Annual Frequency Control Symposium*, 160 (May 1974).
9. Oscillator Model 11205, Austron, Inc., 1915 Kramer Lane, Austin, Texas 78758.
10. Paul H. Carr, "New Temperature Compensated Materials with High Piezoelectric Coupling," *Proc. 30th Annual Frequency Control Symposium*, 129 (June 1976); R.W. Weinert and T.J. Isaacs, "New Piezoelectric Materials which Exhibit Temperature Stability for Surface Waves," *Proc. 29th Annual Frequency Control Symposium*, 139 (May 1975).
11. T.E. Parker and H. Wichansky, "Material Parameters of the Temperature Stable $\text{SiO}_2 / \text{YZ LiTaO}_3$ Structure," *Proc. 1975 IEEE Ultrasonics Symp.*, 503 (Sept. 1975).
12. Some of the data on these devices has previously been reported by T.E. Parker and M.B. Schulz, "Stability of SAW Controlled Oscillators," *Proc. 1975 IEEE Ultrasonics Symposium*, 261 (Sept. 1975).
13. J.R. Vig, J.W. LeBus and R.L. Filler, "Further Studies on U.V. Cleaning and Ni Electrobonding," *Proc. 29th Annual Frequency Control Symposium*, 220 (May 1975); J.R. Vig, C.F. Cook, Jr.,

K. Schwidtal, J.W. LeBus, and E. Hafner, "Surface Studies for Quartz Resonators," *Proc. 28th Annual Frequency Control Symposium*, 96 (May 1974).

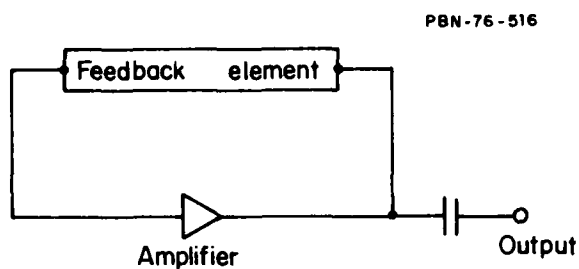


Figure 1 Basic Stabilized Oscillator. Stability is determined primarily by the electrical parameters of the feedback element.

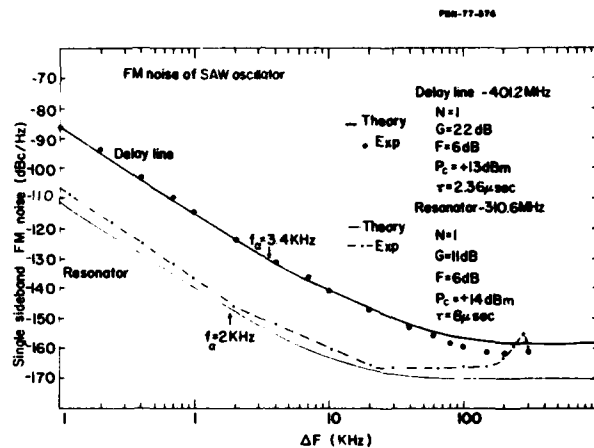


Figure 3 Measured and Calculated FM Noise for a Delay Line and a Resonator.

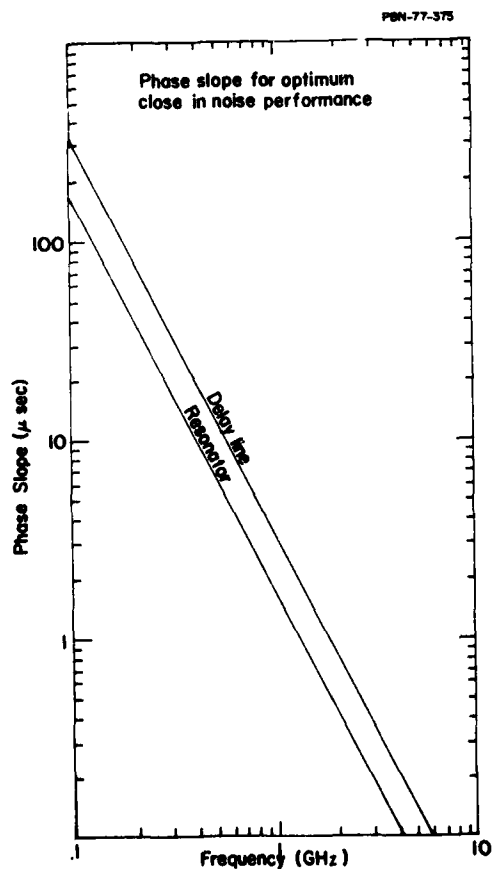


Figure 2 Optimum phase slope as a function of frequency for delay lines and resonators on ST-cut quartz.

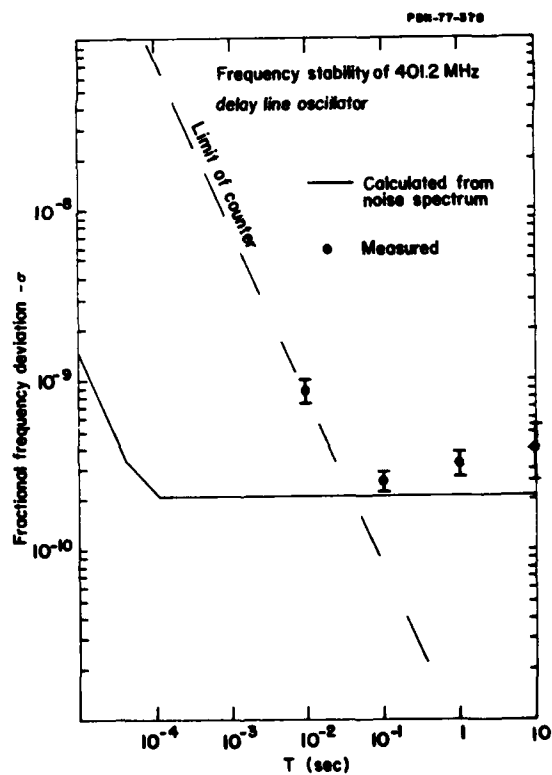


Figure 4 Fractional Frequency Deviation of the Delay Line Shown in Fig. 3.

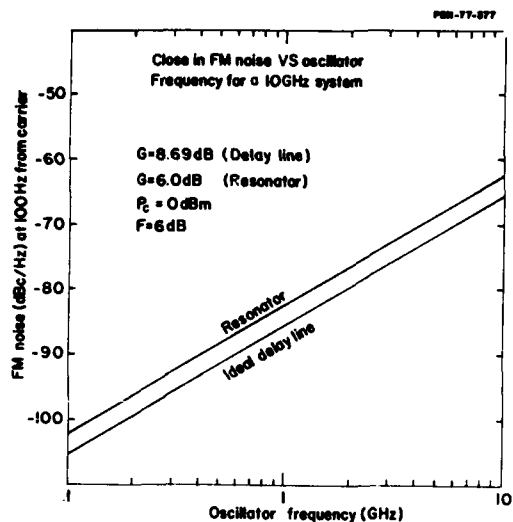


Figure 5 Close in FM Noise of a 10-GHz System as a Function of Oscillator Frequency.

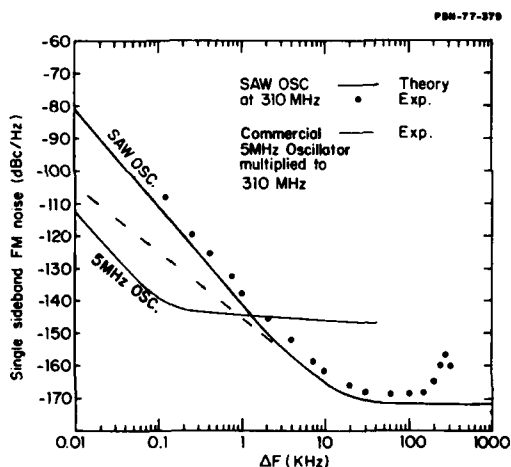


Figure 6 Comparison of FM Noise Between a Commercial 5-MHz Oscillator Multiplied to 310 MHz and a 310 MHz SAW Resonator.

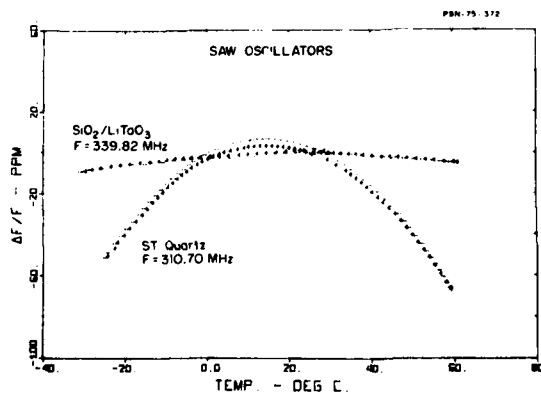


Figure 7 Temperature Dependence of an $\text{SiO}_2/\text{LiTaO}_3$ Oscillator Compared to an ST-Quartz Oscillator.

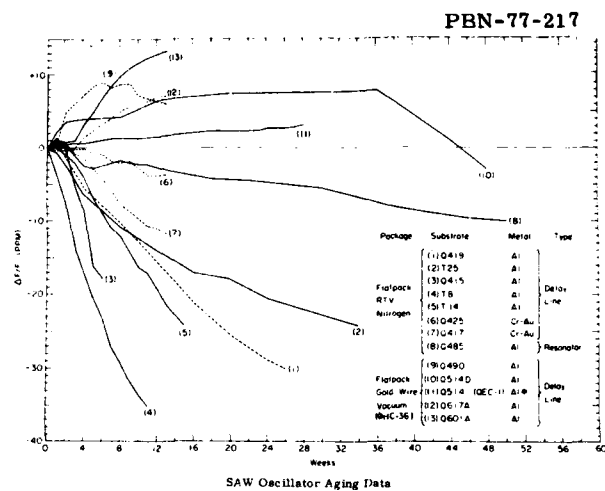


Figure 8 Aging Characteristics of SAW Stabilized Oscillators.

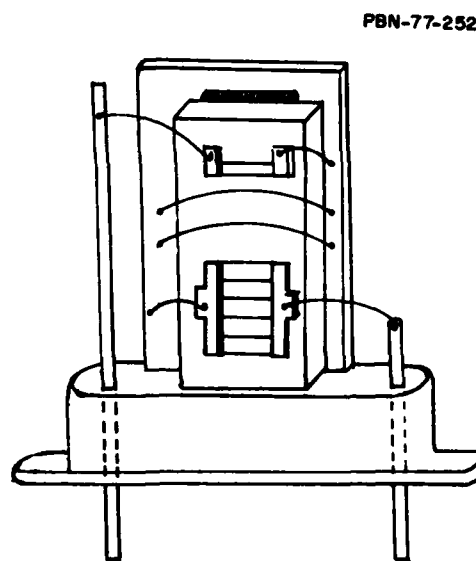


Figure 9 Type HC-36/u Holder Showing SAW Mounting Arrangement.

SAW OSCILLATORS FOR PHASE LOCKED APPLICATIONS

Thomas R. Joseph
Hughes Aircraft Company, Fullerton, CA 92634

Abstract

Many system applications require coherent frequency sources in the UHF frequency range. These sources can be realized by SAW oscillators which are phase locked to a low frequency system reference. The SAW oscillator may be implemented as either a delay line or a resonator controlled oscillator. The choice depends upon a number of factors, including oscillator noise characteristics, temperature effects, circuit complexity (cost), and ease of phase locking. This paper describes the results of an experimental comparison between delay line and resonator oscillators designed for use as phase locked frequency sources in the UHF frequency range. Phase noise measurements for each type of oscillator is presented and compared with the noise characteristics of a contemporary crystal oscillator.

Results of this study show that simplifying modifications can be made in the design of conventional delay line oscillators when they are used in a phase locked mode, and that the delay line must suppress oscillation primarily at the unwanted harmonics. The use of a two-part phase shifter to compensate for variations on the acoustic phase of the delay lines from unit to unit is discussed.

Introduction

Many current system applications require coherent frequency sources in the UHF frequency range. These sources can be realized by surface acoustic wave, SAW, oscillators which are phase locked to a low frequency system reference. Figure 1 shows a block diagram for a phase locked oscillator. It consists of a voltage controlled oscillator, VCO, which operates in the frequency range of interest, a local oscillator, LO, (system reference) which may be much lower in frequency, to which the VCO is to be locked and a phase detector which provides an output signal proportional to the phase difference between the VCO and the LO.

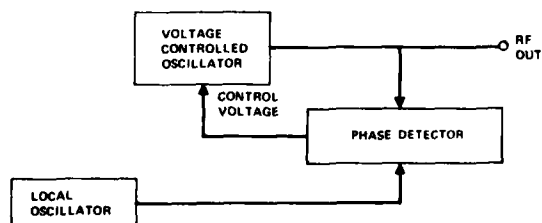


Figure 1. A Phase Locked Oscillator

The three types of VCO to be discussed in this paper are the SAW delay line (DLO), SAW two-port resonator, (2-PRO), and SAW one-port resonator (1-PRO) oscillators. The choice of one of these three oscillator types for a particular application is dependent upon several factors including, long term drift (aging), medium (temperature variation) and short term (noise) stability, circuit complexity, ease of device fabrication and frequency settability.

It is the purpose of this paper to discuss the results of a comparative study of the DLO and the 1-PRO, which were considered for use as phase locked frequency sources in an experimental communications system and the results of a study of 2-PRO's which operated at similar frequencies.

As many of the design considerations for the free running VCO and many of its properties determine the overall behavior of the phase locked oscillator; much of the following discussion will be applicable to SAW oscillators in general. However, those properties which are peculiar to the phase locked environment will be emphasized. Before comparing these three oscillator types an individual description of each will be given. The DLO is considered first and it is seen that the constraints on the delay line design^{1,2} are somewhat different when the DLO is used in a phase locked application than when used as a stand alone oscillator.

The Delay Line Oscillator

The DLO, as shown in Figure 2, consists of an amplifier connected to a feedback loop containing a SAW delay line and two phase shifters, one fixed or static (φ_s) and one variable or dynamic ($\varphi_D(V)$).

The conditions for oscillation are:

$$A L P = 1 \quad (1a)$$

$$\theta_A + \theta_L + \varphi = 2m\pi \quad (1b)$$

Where A, L and P are the gains and θ_A , θ_L , and φ are the phases of the amplifier, delay line and phase shifters respectively, ($\varphi = \varphi_s + \varphi_D(V)$), and m is an integer.

The phase condition (1b) determines the frequencies of oscillation while the frequency response of the delay line, using (1a), suppresses oscillation at all but one of the possible oscillation frequencies. Frequency variation is obtained by varying the input voltage, V, to the dynamic phase shifter, $\varphi_D(V)$.

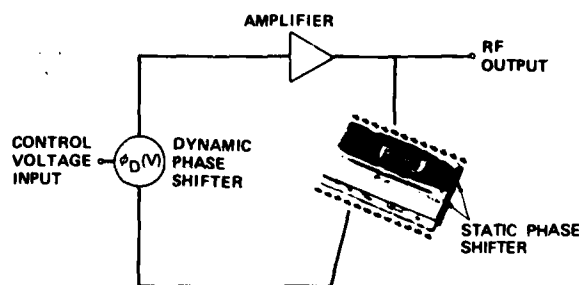


Figure 2. Delay Line Oscillator Showing Stripline/Microstrip Phase Shifter

The dominant term in (1b) is $\theta_L = 2\pi f\tau$, where τ is the time delay of the delay line. Thus, the possible oscillation frequencies, f_m , are given by

$$f_m = \frac{1}{\tau} \left[m - \frac{(\theta_A + \varphi_s + \varphi_D(V))}{2\pi} \right] \quad (2)$$

and the spacing between adjacent oscillation frequencies is

$$\Delta f = f_{m+1} - f_m = 1/\tau \quad (3)$$

As was indicated above suppression of oscillation at all but the desired frequency, f_0 , is accomplished by tailoring the delay line amplitude response so that its transmission peaks only at f_0 . The insertion loss at the other values of f_m must be large enough so that (1a) is not satisfied. This is usually accomplished using a geometry described by Lewis,¹ consisting of a broadband input transducer with a thinned narrowband transducer (tapped array) at the output. Under ideal conditions this provides suppression of all but the desired modes. However, due to fabrication errors, mass loading, and circuit phase effects, (time delay or phase shift resulting from the transducer) this ideal mode suppression is rarely realized in practice. We have found, however, that imperfect mode suppression is adequate if the adjacent modes are at least 10 dB down relative to the insertion loss at f_0 .

In a phase locked implementation the constraints on the oscillator change. Mode suppression is most critical at frequencies which correspond to the possible lock frequencies, f_L , which are integral multiples of the system reference frequency,

$$f_L = l f_R$$

where l is an integer and f_R is the system reference. This suppression can be achieved using a Lewis type delay line but it can also be realized using two periodic transducers whose null bandwidth is $2 f_R$. Consequently, the time length of each transducer is

$$\tau_{IDT} = 1/f_R$$

For the Lewis design the time delay, τ , of the delay line is equal to τ_{IDT} of the narrowband transducer. However, we have found in practice that 10 dB of mode suppression is sufficient to suppress unwanted modes. Thus, it is possible to increase τ for a delay line consisting of two uniform periodic transducers up to a maximum of $1.8 \tau_{IDT}$. This can be seen by referring to Figure 3 in which the normalized insertion loss of an ideal delay line with two identical transducers of length τ_{IDT} is plotted. The 10 dB line (shown dotted) intercepts the main lobe at frequencies whose separation is $2\Delta f$ is related to f_R by

$$f_R/\Delta f \approx 1.8$$

using (3) to substitute for Δf gives the result $\tau \approx 1.8 \tau_{IDT}$. In a phase locked oscillator it is necessary that $\tau \approx 1/f_R$. This prevents the adjacent lock frequencies from coinciding with possible values of f_m . The mode suppression constraint can be relaxed as oscillation at an adjacent mode frequency will not null the output of the phase detector and

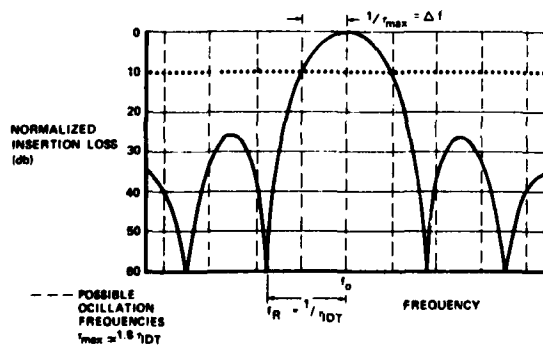


Figure 3. The Relation Between τ_{IDT} and τ_{max}

the lock circuitry will pull the oscillator until the loop mode hops to the desired frequency. Some care must be taken, if the phase locking circuitry is required to aid in the mode suppression, to insure that the phase variation under voltage control is sufficient to drive the loop to a frequency where it can no longer oscillate. Otherwise, the loop can, as indicated by Browning, et. al.,³ remain oscillating on the unwanted mode and the oscillator will not achieve phase lock.

Delay lines using both the Lewis type design and the identical transducer design just discussed, were constructed during this study. Figure 4 shows the measured insertion loss and the insertion phase (θ_L) of two delay lines with identical transducers operating at 313.75 and 393.75 MHz. $\tau_{IDT} = 0.833$ microseconds was chosen. This is slightly larger than the theoretical value to insure the suppression of the adjacent lock frequencies despite variations in temperature. The transducers had double electrodes and $\tau = 1.1$ microsecond or $1.3 \tau_{IDT}$. The loop mode frequencies occur at the frequencies where $\theta_L(f) = \theta_L(f_0)$. At those frequencies the insertion loss is substantially greater than 10 dB and the suppression level is at least 40 dB at the lock frequencies.

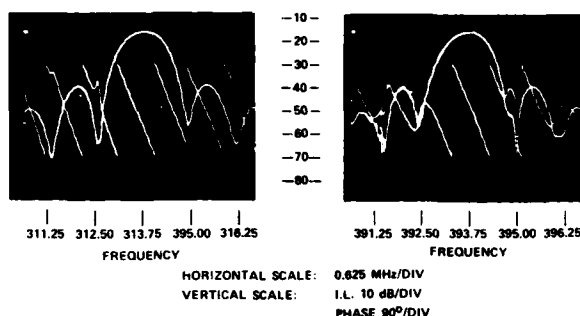


Figure 4. Measured Insertion Loss of Two Oscillator Delay Lines Using Identical Uniform Periodic Transducers

While the delay line phase is the largest term in Eq. (1b), ϕ is critical in determining the oscillation frequency. The function of the phase shifter is to provide a voltage dependent variation in the loop phase enabling the oscillator to be phase locked. However, this is really two separate processes. The first is to pull the oscillator up to the desired lock frequency by compensating for transducer circuit effects, fabrication defects and any other permanent perturbation in the loop phase, and the second is to compensate for the phase shifts due to variations of ambient temperature, the affect the delay line, the amplifier and the phase shifter itself, and shifts due to delay line aging. As such, the phase shifter can be divided into two parts, these were shown in Figure 2 as a static phase ϕ_s and a dynamic one $\phi_D(V)$. The dynamic shifter must compensate for all of the time dependent phase shifts in the loop consequently its maximum phase variation, $\Delta\phi_{Dmax}$, must be greater than the expected time dependent phase shift in the loop in order to maintain phase lock. The changes in θ_L due to aging and temperature variation are usually the largest contribution to changes in the total loop phase, however, temperature variations of the amplifier and of the phase shifter itself must also be considered. The variation of the frequency due to changes in the elements of the oscillation loop is obtained by taking differentials of Eq. (2) and using (1b).

$$\frac{\Delta f_m}{f_m} = \frac{1}{\theta_L} (\Delta\theta_L + \Delta\theta_A + \Delta\phi) \quad (5)$$

$\Delta f_m/f_m = 0$ in a phase locked system and consequently,

$$\Delta\phi_D = -(\Delta\theta_L + \Delta\theta_A + \Delta\phi_s) \quad (6)$$

Since the two circuit terms $\Delta\theta_A$ and $\Delta\phi_S$ are expected to be small (6) reduces to the condition

$$\Delta\phi_{Dmax} \approx \left| \Delta\theta_L \right| \quad (7)$$

The LHS of Eq. (7) is usually fixed by the design of the phase shifter. Voltage variable shifters with $\Delta\phi_{Dmax} \approx \pi$ are difficult to design with a flat amplitude response, and good temperature stability. Consequently, it is desirable to limit the amount of phase shift required as much as possible. Eq. (7) gives a relation between $\Delta\phi_{Dmax}$ and the length of the delay line. Rearranging Eq. (7) gives a relation between $\Delta\phi_{Dmax}$ and N the length of the delay line in wavelengths.

$$N \leq \Delta\phi_{Dmax} / 2\pi \left| \frac{\Delta\theta_L}{\theta_L} \right| \quad (8)$$

Figure 5 is a plot of N vs $\Delta\theta_L/\theta_L$ with $\Delta\phi_{Dmax}$ in degrees as a parameter. A limit on delay line length resulting from the expected variation of θ_L with temperature and the design value of $\Delta\phi_{Dmax}$ can be obtained from Figure 5.

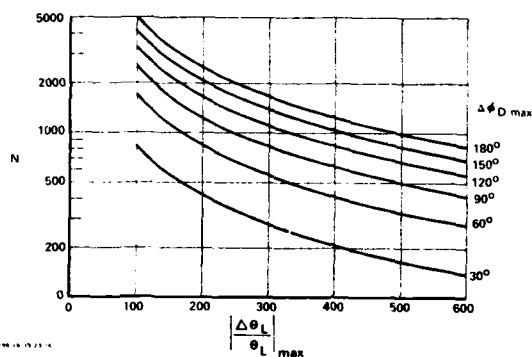


Figure 5. The Relation Between Delay Line Length and Fractional Phase Change for Several Values of $\Delta\phi_{Dmax}$

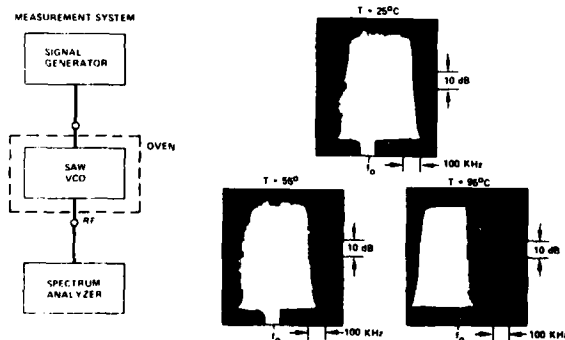


Figure 6. Change in FM Range of a Delay Line Oscillator Over Temperature

Some care must be taken to insure that the range of total loop phase values required to maintain phase lock falls within the range of phase values covered by the dynamic phase shifter. This situation is analogous to the biasing of a transistor at a specific operating point. Figure 6 shows the pulling range of a DLO for V between 0 and 10 volts at $T = 25, 55$ and 95°C . The loop phase has been adjusted so that despite changes in temperature, f_0 still falls within the pulling range of the oscillator. This is accomplished by setting ϕ_S to a value which gives oscillation at f_0 at between 50 and 100 kHz above the low frequency end of the pulling range. If no static phase shifter is used then the phase of

the delay line must be controlled with sufficient accuracy to obtain this same pulling range characteristic. This requires delay line reproducibility of ± 80 ppm. This is not currently practical as metallization thickness and crystal orientation inaccuracies will in general exceed this tolerance. Setting the loop phase to the proper value is accomplished by including a static phase shifter. We implemented this phase shifter by a combination strip line/microstrip meander line structure which is shown in Figure 2. The phase was varied by selectively making and breaking interconnections between the sections of the meanderline. Note, the static shifter must be isolated from the dynamic phase shifter, otherwise the reactive load of the static shifter and transducer combination, which changes as the phase shift is varied, can cause the dynamic phase shifter's characteristics to vary, resulting in changes in the FM characteristic of the VCO.

THE SAW TWO-PORT RESONATOR OSCILLATOR

The SAW two-port resonator⁴ (2-PR) is in many respects similar to the delay line. It is essentially a delay line whose characteristics are modified by the presence of two reflective gratings which surround the transducers of the delay line. As such its use in an oscillator is similar to the DLO. These are however three major differences. 1) A high Q 2-PR is extremely narrowband, 2) the insertion phase slope while constant outside the passband becomes quite steep over the operating bandwidth and 3) because of the energy confining effects of the grating the insertion loss can be potentially 0 dB. Figure 7 shows the insertion loss of a typical 2-PR fabricated on ST quartz using shorted metal electrode gratings. The narrowband response (3 dB bandwidth = 46 kHz) and the steep phase slope can be easily seen. This 2-PR has a loaded Q of 8600 at 394 MHz.

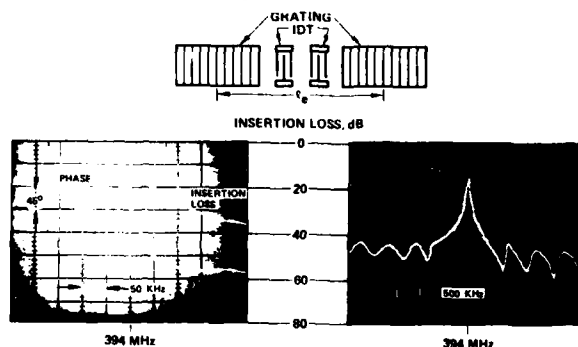


Figure 7. Measured Response of a Two-Port SAW Resonator Fabricated on ST Quartz Using Shorted Metal Electrode Gratings

The low insertion loss and high Q of the two-port resonator make it potentially useful as a stable stand alone frequency source. However, its extremely narrow bandwidth makes it difficult to fabricate devices which oscillate at the phase lock frequency f_0 and which can hold that frequency over temperature. For example, Figure 8 shows the measured insertion loss of the delay line used in this study compared to the response of the 2-PR resonator. Note the delay line has a Q of 2500 and a 3 dB bandwidth of 1.3 MHz. The 2-PR has a Q of 8600 and a 3 dB bandwidth of 46 kHz. Even with a Q of 2500 the 3 dB bandwidth would only be 160 kHz, and this lower Q value can be achieved only at the expense of increased insertion loss.

The application for which the phase locked oscillators were being studied did not allow the use of an oven to temperature stabilize the SAW device. Thus the question of the applicability of the 2-PRO as the VCO in a phase locked oscillator depends upon the ability to hold oscillation at f_0 over the temperature. Our two-port oscillators were found to drift by 140 kHz over the 0 - 100°C range specified for

this application. As this is greater than their 46 kHz bandwidth they would not be suitable without some improvement in the temperature stability. The question of temperature stability will be discussed further below.

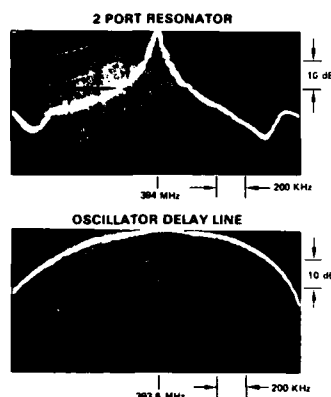


Figure 8. Comparison of Passband Width Between the 2-Port Resonator and an Oscillator Delay Line

THE SAW ONE-PORT RESONATOR OSCILLATOR

An alternative to the delay line oscillator is the SAW one-port resonator oscillator. The one-port resonator⁵ (1-PR) is analogous to the bulk crystal resonator and can be used to build oscillators in a similar manner. The 1-PR has the same narrowband nature of the 2-PR however, through the use of spreading coils the pulling range of the oscillator can be increased.

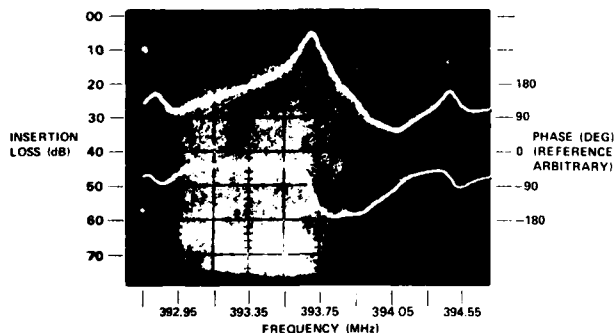


Figure 9. Series Transmission Measurement of a One-Port Resonator Which is Shunted by a Coil to Resonate the Static Capacitance of the Transducer

One means of determining the properties of a 1-PR is to measure its frequency response as a series transmission element in a 50 Ω system. This is a particularly appropriate means of measuring the properties of the resonator as it is often used as a series feedback path in many oscillator designs. Figure 9 shows the measured response of a SAW resonator which is shunted by a coil to resonate out the effects of both stray capacitance and the static capacitance of the transducer. This plot indicates the major differences between this type of device and the delay line. The minimum insertion loss is 5 dB instead of the 15 - 30 dB usually seen in delay lines, and the insertion phase is not linear with frequency except over several separate frequency regions. Two of these regions are of particular interest for the oscillator designer. The region about series resonance is the region where the steepest phase slope occurs and the transmission peaks. The resonant Q measured from the slope of the phase is 4650. This region which is 200 kHz wide is the region over which the oscillator would have the best short term stability, due to the high Q, and low insertion loss. A properly designed stand alone oscillator should operate in this region. In a phase locked oscillator the stability of the reference oscillator determines the stability

of the VCO. Thus the requirement that the oscillator be capable of maintaining the phase lock frequency over temperature is the most critical, and the oscillator may be operated in the region below resonance where a 400-500 kHz pulling range is obtainable. A resonator was placed in a standard Clapp type transistor oscillator circuit. The circuit is shown in Figure 10. The frequency is varied by changing the varactor control voltage, V. Buffer amplifiers were used to isolate the oscillator. The final oscillator constructed had two outputs each with a 7 dBm signal level and 80 dB of suppression of the system reference and its modulation products.

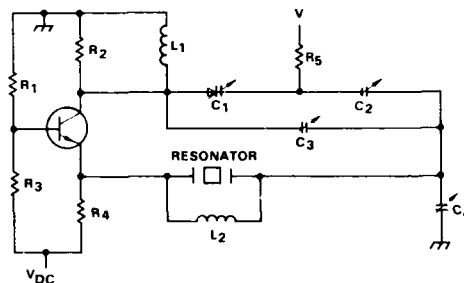


Figure 10. SAW Resonator Oscillator Circuit Diagram

COMPARISON OF OSCILLATOR TYPES

The decision as to which type of oscillator to use depends upon a number of factors. These include sensitivity to variations in fabrication, ease of fabrication, circuit complexity, the noise characteristics of the free-running oscillator and, in particular, for a phase-locked source, temperature variation of frequency and the ability to maintain phase lock despite variations in temperature and long term drift. The main results of this study were to shed some light on these factors.

Fabrication

Several factors in the fabrication process were found to impact on the delay line performance. Pattern defects consisting of broken fingers did not appear to change the insertion loss characteristic although shorts between electrodes which were large compared to a finger width rendered the device inoperative. Control of metal thickness to within 100 \AA of the nominal value was found to be necessary to insure that the measured center frequency corresponded to the design value. Table 1 gives measured values of f_0 for delay lines with differing metal thicknesses. As can be seen the center frequency changed by approximately 200 kHz when the thickness was changed by 300 \AA . This is a much greater shift than is predicted by theory. The 1650 \AA devices were built using a different photomask. The 313 MHz delay line was found to have a stripe to gap ratio less than 1:1 which caused the frequency to be the same as seen for a 1500 \AA metalization. The 1500 \AA devices had a stripe to gap ratio close to 1:1.

Table 1

Metal Thickness (\AA)	f (313.75) (MHz)	f (393.75) (MHz)
1500	313.96	393.92
1650*	313.97	393.66
1800	313.57	393.55

*A new mask was used for the 1650 \AA devices.

Initially by chemically etching the 1800 Å thick metalizations lines with the correct center frequency were obtained. Currently the 1650 Å devices which are as much as 200 kHz high have been used. This is possible because the center frequency tends to decrease with temperature allowing the DLO to hold phase lock. Delay lines with center frequencies which are low are not useable.

The sensitivity to shorts of the resonator geometry is limited to the transducer regions which comprise only a small fraction of the device geometry. The gratings which consist of shorted metal electrodes are relatively immune to such defects. However, the two resonators are narrow band devices and metal thickness and stripe to gap variations can cause shifts in the resonant frequency. For example, the resonant frequencies of a group two-port resonators which were fabricated on the same substrate were found to vary by as much as 65 kHz. This variation is 1.3 times the 3 dB bandwidth of the 2-PR's.

The delay line presents a somewhat greater problem in photolithography. In order to suppress electrode reflections which can seriously distort the passband shape, double electrode transducers are necessary. Double electrodes require greater line width resolution for pattern definition. This increased resolution requirement has not been a serious problem at 393 MHz, but it does render the device more difficult to fabricate and increases the incidence of fabrication defects. The resonators however, are essentially single electrode structures and the line widths are twice that of the delay line.

Temperature Effects

The frequency of a phase locked oscillator does not vary with temperature as long as the locking circuitry is capable of compensating for the changes in the SAW element and in the auxiliary circuit. While the dominant variation with temperature, resulted from phase changed in the SAW element, the gain of the amplifiers and dynamic phase shifters was seen to decrease. This resulted in a decrease in the pulling range of the oscillator. This can be seen in Figure 6 where at 25°C the fm range is 340 kHz while at 95°C this has been reduced to 250 kHz.

Both the DLO and the 1-PRO had sufficient fm capability to maintain phase lock from 0 to 100°C. However, the 2-PRO did not. The 2-PRO pulls 140 kHz over the temperature range. If the lock frequency is 70 kHz above f_0 at T_0 then at $T = T_0$ and at $T = 100^\circ$ where the passband will have shifted so that f_0 is 70 kHz below center frequency. Then the insertion loss will have decreased by 12 dB. Unless this much excess gain is included in the loop the oscillator will not hold phase lock over temperature.

The change in acoustic phase with frequency results from the variation of SAW velocity with temperature which is a property of the choice of crystal, crystal orientation and the fractional surface metalization. The SAW devices used in this study were fabricated on rotated Y-cut quartz. ST-cut quartz was used for the delay lines and the 2 PR's and a 5° rotation from ST was used for the 1 PR's. The measured zero slope temperature, T_0 , and second order temperature coefficient of frequency, C_2 , are given in Table II, for these devices.

T_0 is reduced in these devices because of large fractional surface metalization.⁷ The temperature characteristics of each device did not depend upon the type of device but instead upon the physical parameters of metal thickness and crystal orientation. With careful choice of that orientation $T_0 = 50$ can be realized for each device type thus minimizing the phase variation. With $T_0 = 50$ the drift of the 2 PRO from 0 - 100°C would not exceed its 3 dB bandwidth.

Table II.

Device Type	Orientation*	T_0 (°C)	C_2 ppm/(°C) ²	$\left \frac{\Delta f}{f} \right _{\max}$ (0-100°C)
Delay Line	ST (42.75°)	-18.6	-.035	492 ppm
1-port Resonator	37.7°	31.2	-.032	151 ppm
2-port Resonator	ST (42.75°)	-12.9	-.028	357 ppm

*Orientation given in the notation of Ref. 8.

Short Term Stability

Measurements were made of the short term stability of each of these three oscillator types. The measurement was made using the apparatus shown on the left of Figure 11. For this measurement two oscillators are required. One is configured as a fixed oscillator while the second is a voltage variable oscillator which is phase locked to the first oscillator. A quadrature mixer is used to beat the two signals together 90° out of phase. The resulting signal is uncorrelated noise which is then fed into a low frequency spectrum analyzer.

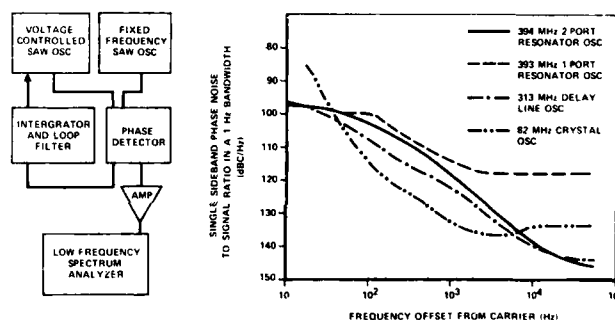


Figure 11. Measured SSB Phase Noise for Three Types of SAW Oscillator. The response for a crystal oscillator is shown for comparison. The data has been normalized to 393 MHz.

Figure 11 shows the measured single sideband phase noise to signal ratio (dBc in a 1 Hz bandwidth) of SAW 393 MHz one-port and two-port resonator oscillators, a 313 MHz delay line oscillator and a standard 82 MHz crystal oscillator. The noise spectra of the DLO and the crystal oscillator has been normalized to 393 MHz to give a more meaningful comparison.

As can be seen in Figure 11 all four oscillator types show comparable noise spectra out to about 2000 Hz. Beyond this frequency the 1-PRO joins a noise floor at -116 dBc while the other three spectra continue to decrease till they join noise floors at 134, 143 and 145 dBc for the crystal oscillator, DLO and the 2-PRO respectively.

According to the theory of Lewis¹ the noise characteristic of a DLO (and also the 2-PRO) is improved by using a higher Q, lower insertion loss element. Thus the 2-PRO because it has both of these properties has the potential to produce a superior noise characteristic. However, in Figure 11, the 2-PRO noise characteristic is seen to be 2-4 dB higher than that of the delay line. The reason the delay line characteristic is lower is not fully understood at present. One problem which was encountered in making this measurement was the difficulty of obtaining two devices with the same center frequency. The two devices used had center frequencies which differed by 10 kHz and the

measurement was made with both oscillators tuned to a frequency in between. This results in a small increase in the insertion loss of each device.

The theory of Leeson⁹ for the crystal oscillator which can be applied to the 1-PRO indicates that the phase noise spectrum, $S_{\phi}(f)$, as

$$S_{\phi}(f) = \left[\frac{\alpha}{f} + 2 \frac{F K T}{P_s} \right] \left[1 + \left(\frac{f_o}{2Qf} \right)^2 \right] \quad (9)$$

where α is a constant dependent upon the choice of amplifier, F is the noise figure, K , Boltzmann's constant, P_s is the power level at the amplifier input, and f_m is the frequency offset from the carrier. From this it is seen the noise floor of the two-sided spectrum is $2 F K T / P_s$ and the $1/f^2$ portion of the curve joins this noise floor at $f = f_o/2Q$. Note, Q is the loaded Q of the resonator in the oscillator circuit.

The frequency at which the resonator oscillator spectrum joins the noise floor indicates, using Eq. (9), the loaded Q for the feedback path including the SAW device is 79,000. This value is about 10 times the Q of the resonator. Thus, it appears that there is some internal or external noise source which has increased the noise floor and is masking the actual oscillator characteristics.

Each of these three oscillator types has been shown to be capable of use as a low noise frequency source. However, both of the resonator oscillators require further effort to fully realize their capability for good short term stability.

CIRCUIT COMPLEXITY

The two factors which impact on the complexity of the circuitry are the insertion loss of the SAW element and the method by which the insertion phase of the SAW device is adjusted to maintain oscillation at the proper frequency. The two resonators because of their inherently lower insertion loss required fewer stages of gain in the feedback path. The 1-PRO circuit as shown in Figure 10 consisted of only a single stage transistor amplifier. In contrast three 9 dB modular amplifiers were needed to overcome the high insertion loss of the delay line.

The inability to fabricate the delay lines with sufficiently small phase tolerances required the use of a static phase shifter. This shifter requires extra circuit board space and must be set individually for each device. The two-port resonator because it is essentially a narrowband delay line would require a similar type of phase shifter arrangement. Adjustment of the phase shifter, however, can only be effective if the desired frequency falls within the resonator passband. The 1-PR does not require this type of adjustment. It functions as a reactive element whose impedance takes on a limited range of values near the acoustic resonance. Consequently, the frequency adjustment, assuming the SAW device is resonant at the correct frequency is much simpler to implement.

Conclusion

As a stable stand alone frequency source the two resonator oscillators have the greatest potential because of their higher Q and lower insertion loss. However, for a phase locked application the two-port resonator is not suitable. The one-port resonator is similarly unsuitable if operation in its narrowband high Q region is required.

However if a lower Q element is acceptable a 1-PRO can provide a pulling range sufficient to be useful as a phase locked source. Comparison of the one-port resonator and the delay line indicates they are comparable in physical size but the static phase shifter required by the delay line necessitates utilization of more circuit card area and individual adjustment of each device. The higher insertion loss of the delay line also requires greater amplifier gain.

Thus while both the delay line and the one-port resonator oscillators were found to be capable of satisfying the requirements of the phase locked implementation under consideration, the SAW resonator oscillator is felt to be the more advantageous approach.

Acknowledgements

The author wishes to express his gratitude to C. Stout and F. Kandra for their technical assistance, and to B. Dobratz and M. Ammon for performing the noise measurements.

References

1. M. F. Lewis, "The Surface Acoustic Wave Oscillator - A Natural and Timely Development of the Quartz Crystal Oscillator," *Proceedings of the 28th Annual Symposium on Frequency Control*, U.S. Army Electronics Command, Fort Monmouth, N.J., 1974, pp. 304-314.
2. H. G. Vollers and J. T. Claiborne, "RF Oscillator Control Utilizing Surface Wave Delay Lines," *Proceedings of the 28th Annual Symposium on Frequency Control*, U.S. Army Electronics Command, Fort Monmouth, N.J., 1974, pp. 256-258.
3. I. Browning, J. G. Crabb, and M. F. Lewis, "A SAW Frequency Synthesizer," *1975 Ultrasonics Symposium Proceedings*, IEEE Cat. No. 75 CHO 994-4SU, pp. 245-247.
4. T. R. Joseph and K. M. Lakdn, "Two-Port Cavity Low-Insertion-Loss Delay Line," *Appl. Phys. Lett.* 26, 7, pp. 364-365. (4/75)
5. There are several papers on the subject in the 1974 Ultrasonics Symposium Proceedings, IEEE Cat. No. 74 CII0 896-1SU, pp. 245-267.
6. D. Penunuri and K. M. Lakdn, "Surface Acoustic Wave Velocities of Isotropic Metal Films on Selected Cuts of Bi₂GeO₁₀, Quartz, Al₂O₃, and LiNbO₃," *IEEE Trans. on Sonics and Ultrasonics*, SU-21, 4, pp. 293-295. (10/74)
7. S. J. Kerbel, "Design of Harmonic Surface Acoustic Wave (SAW) Oscillators Without External Filtering, and New Data on the Temperature Coefficient of Quartz," *1974 Ultrasonics Symposium Proceedings*, IEEE Cat. No. 74 CII0 896-1SU, pp. 276-281.
8. J. F. Dias, H. E. Karrar, J. A. Kusters, J. H. Matsinger, and M. B. Schulz, "The Temperature Coefficient of Delay-Time for X Propagating Acoustic Surface-Waves on Rotated Y-Cuts on Quartz," *IEEE Trans. on Sonics and Ultrasonics*, SU-22, 1, pp. 46-50, (1/75).
9. D. B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum," *IEEE Proceedings (Letters)*, 54, 2, pp. 329-330. (2/66)

300 MHz OSCILLATORS USING SAW RESONATORS AND DELAY LINES

E. J. Staples and T. C. Lim
Science Center, Rockwell International
Thousand Oaks, California 91360

Summary

The design and performance of fundamental mode crystal oscillators utilizing SAW resonators and delay lines at 300 MHz is presented. Emphasis is on the measurement of phase noise close to the carrier (1 Hz to 10 kHz), and it is shown that the SAW resonator oscillator possesses lower phase noise than comparable delay line oscillators. The lower phase noise of these oscillators is a direct consequence of the higher Q or phase slope achieved in the feedback network of the oscillator circuit. Experimental measurements of frequency domain phase noise $S_{\phi}(f)$ and time domain stability $\sigma_y(2, \tau)$ for SAW delay line and resonator oscillators will be compared and discussed.

Introduction

An important new circuit application to evolve from present surface wave technology is in the area of oscillator frequency control. Recent literature has shown that a surface wave delay line can be used as the feedback element of an RF oscillator for generation of moderate to wide bandwidth frequency-modulated signals in the VHF/UHF range. The surface wave delay lines used to stabilize the oscillator's frequency can be designed to have Q's in the range between those of crystal and LC resonators used in other oscillator circuits. The frequency stability that can be achieved with the delay line oscillator is limited by the delay line phase slope or Q. As will be shown in this paper, this limitation can be reduced by the use of a SAW resonator in the feedback loop of the oscillator. In this case, the stability of the oscillator depends on the Q of the SAW resonator which can be made considerably higher than delay lines, i.e., typically 30,000. An important advantage of SAW resonator-controlled oscillators is that they can be fabricated at frequencies as high as 1 GHz; hence, stable oscillators can be built for fundamental frequency operation well into the UHF range. This eliminates the need for a complex frequency multiplier chain and greatly reduces the harmonic content and phase noise of the output signal.

The purpose of this paper is to present the results of a recent survey of SAW oscillator performance made at the Rockwell Science Center. The relative stability of SAW oscillators compared to the bulk oscillators has until recently not been well known partially through a lack of commercially-available measurement equipment and also because of the discovery only recently of high Q SAW resonance phenomena itself. In this paper stability is defined in terms of time and frequency domain measurements performed on the HP 5390 stability analyzer. Both contain the same information but viewed through a different perspective. The relative stability of SAW oscillators to that of bulk oscillators is now available as a benchmark for future researchers and system designers. At the present time SAW oscillators are not as stable as bulk oscillators. However, it is possible to improve the SAW oscillator by careful fabrication techniques.

Frequency Stability

The stability of an oscillator is characterized by statistical fluctuations about the average frequency

f . The fluctuations are given in the time domain by $y(t)$ the fractional frequency stability.

$$\frac{\Delta f}{f} = \sqrt{\sigma^2(N, \tau)} \quad (1)$$

where $\sigma^2(N, \tau)$ is the variance of N adjacent samples of the average frequency and τ is the averaging time.

If the oscillation is described by

$$V(t) = V_0 \sin[2\pi f_0 t + \phi(t)] \quad (2)$$

and the fractional frequency offset by

$$y(t) = \frac{\dot{\phi}(t)}{2\pi f_0} \quad (3)$$

the stability may be described in the frequency domain by the spectral density of $y(t)$ which is then related to the spectral density of phase, $S_{\phi}(f)$ or frequency $S_y(f)$ by

$$S_y(f) = \left(\frac{1}{2\pi f_0}\right)^2 S_{\phi}(f) = \left(\frac{f}{f_0}\right)^2 S_{\phi}(f) \quad (4)$$

where f is the frequency offset from f_0 .

The frequency stability of any feedback oscillator depends on the following:

- The stability of the feedback element (resonator) as a function of time (aging), pressure, temperature, drive level, and vibration
- The Q of the feedback element
- Perturbations in the amplifier and feedback element by noise sources

Short-term stability is predominantly influenced by the last two factors. Because of limiting in the oscillator, the exact solution for short-term stability is a complex solution related to a nonlinear system of equations as described by Hafner.¹ Assuming the noise perturbations to be small relative to the total signal allows for certain simplifications to be made and the noise can be considered as a superposition of a number of different components.

Oscillator stability in the time domain is measured by evaluating the Allan variance defined as the mean square value of $\sqrt{1/2} (\bar{y}_k - \bar{y}_{k+1})$:

$$\sigma_y^2(2, \tau) = \left\langle \frac{(\bar{y}_{k+1} - \bar{y}_k)^2}{2} \right\rangle \quad (5)$$

where y_k is the fractional frequency averaged from t_k to t_{k+1} .

AD-A088 221

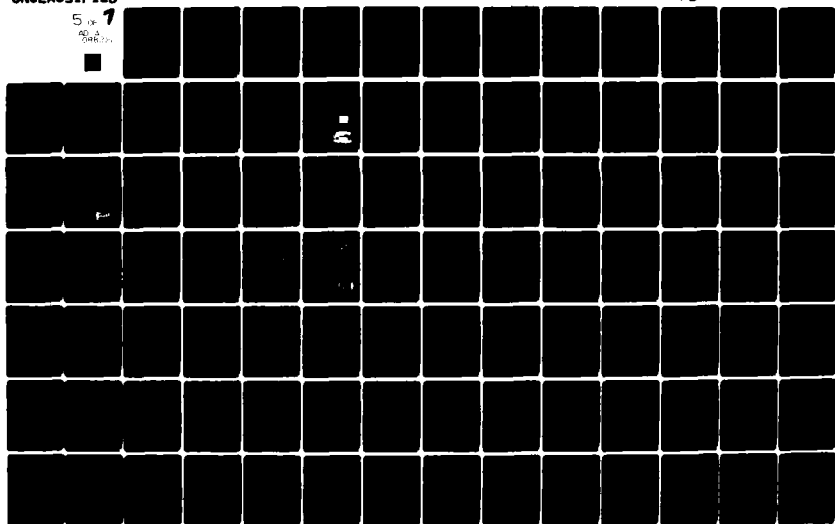
ARMY ELECTRONICS COMMAND FORT MONMOUTH NJ
PROCEEDINGS OF THE ANNUAL FREQUENCY CONTROL SYMPOSIUM (31ST) HE--ETC(U)
1977

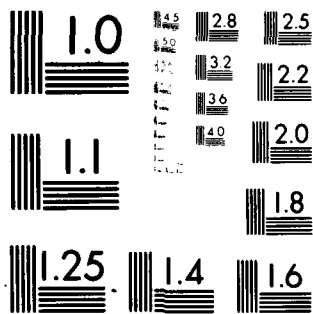
F/G 9/5

UNCLASSIFIED

NL

5 OF 7
AD-A088 221





MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

$$\bar{y}_k = \frac{1}{\tau} \int_{t_k}^{t_{k+\tau}} y(t) dt \quad (6)$$

and $t_{k+1} = t_k + \tau$. The development of frequency counters which can be programmed to make measurements and do arithmetic calculations makes it possible to easily evaluate the variance of frequency fluctuations of any oscillator.

A number of expressions for oscillator noise in the time domain have been derived by various researchers. The predominant components are due to $1/f$ and additive noises. At present, no mathematical expression exists that describes the dependence of flicker noise on oscillator configuration or on parameters of circuit components in the time domain. An expression for the standard deviation of rms fractional frequency fluctuations versus averaging time, derived for an oscillator by E. Hafner² is given by

$$\sigma\left(\frac{\Delta f_{rms}}{f}\right) = \frac{2\pi}{\tau} \sqrt{\frac{4kTF}{P_{osc} Q_L f_o}} + C \quad (7)$$

where

F = effective noise figure of the limiting amplifier in the oscillator feedback loop

kT = -174 dBm/Hz bandwidth at room temperature

P_{osc} = total rf power dissipated in the oscillator resonant circuit

Q_L = operating, or loaded, quality factor of the oscillator resonant circuit

f_o = average frequency of oscillations

C = flicker noise constant

From this expression an estimate of oscillator stability can be made. The important parameter to note is $P_{osc} Q_L f_o$. Bulk crystal oscillators at 10 MHz operate with a loaded Q of typically 500,000 and a power level of -60 dBm. For a SAW oscillator to have the same stability at 100 MHz and the same power level, its Q need only be 50,000. Similarly, at 1000 MHz the Q need only be 5000. These Q 's are close to the intrinsic limit of SAW resonators. Our experience at the Science Center indicates less Q can be tolerated since the SAW oscillator crystal can operate at considerably higher power levels without drive-level effects occurring, i.e., typically -50 to -40 dBm.

Stability Measurements System

At the Science Center an integrated crystal measurement system has recently been assembled. A block diagram of the system is shown in Fig. 1. The heart of the system is the HP 9825 calculator and interface bus. Utilizing the bus, virtually any instrument may "talk" or "listen" to any other instrument. Data may be collected by the calculator, processed, and compared to theory directly on plotters, or printers as available. The emphasis in this paper is on the HP 5390 stability analyzer portion of the measurement system. The analyzer is capable of measuring phase noise as close as .01 Hz to the carrier which can be as high as 18 GHz. The system consists of five major components shown in Fig. 1: a

high-resolution reciprocal counter (HP 5345A) with external gating capabilities; a measurement storage plug-in unit (HP 5358A); a mixer/amplifier unit (HP 10830A); a desk-top calculator (HP 9825A), and a printer/plotter output device (HP 9871A). Communication and control between the various instruments is provided by a digital interface system (HP-interface bus). Measurement data is also sent to the calculator via the interface bus. The calculator is programmed from its keyboard to process the data received from the counter and plug-in. The printer device is used to output the processed results either in numeric or plotted form.

The performance of this system is best explained by the relationship between the Allan variance and spectral density of a system with transfer function $H(f)$:

$$\sigma^2(2, \tau) = \int_0^\infty S_y(f) |H(f)|^2 df \quad (8)$$

The transfer function of a counter performing the Allan variance has a broad peak near $f = 1/2\tau$. Because this peak is so broad, the Allan variance is replaced by the Hadamard variance which can be programmed to have a much narrower transfer function as discussed by Baugh.³ The counter is programmed to measure this variance, the transfer function is then known, and the calculator then processes the data to yield the phase noise of the oscillator being measured based on Eq. (8). The complete theory of this measurement technique has been described by Peregrino and Ricci.⁴

Surface Wave Oscillators--Comparisons

Shown in Fig. 2 is the block diagram of a typical oscillator circuit. It consists of an amplifier having gain "A" and a feedback network with gain "B." These are connected in a closed loop such that positive feedback is provided from the output of the amplifier back to its input. The equation describing the criterion for oscillation is

$$AB = 1$$

where A and B are vector quantities. This equation states that for oscillation to occur, the overall loop gain must be unity and the total phase shift around the loop must be zero or any multiple of 2π .

In the SAW delay line oscillator, the feedback network takes the form of a delay line of time τ_d . The phase slope of the delay line determines to a large extent the phase stability of the oscillator and is given by $d\theta/d\omega = -\tau_d = -N/f$ where N is the number of wavelengths between the input and output transducer phase centers. A typical tapped delay line electrode with $N = 600$ was used in our study and is depicted in Fig. 2. The complete feedback network contained matching networks to reduce the large insertion loss of the delay line and provide the correct phase condition for oscillation. The effective Q of the feedback network is simply πN which for our case was nominally 2000.

In the case of a SAW-resonator oscillator the feedback network may take the form shown in Fig. 3. The SAW resonator is represented as a crystal equivalent circuit. Shown in Fig. 3 is the phase and amplitude response of the feedback network. The low insertion loss and high phase slope are readily apparent. The phase slope of the feedback network is given by

$d\theta/d\omega = -2Q/\omega_0$ where for the resonator

$$Q = \frac{\omega L_1}{R_1}$$

In our experiments Q's of 5000-10,000 were readily obtained using simple aluminum metallizations on rotated Y cuts of quartz. In this same symposium Q's as high as 30,000 were reported by Adams and Kusters⁵ using a plasma etching fabrication technique.

To evaluate SAW oscillators the measurement capabilities of the HP 5390 stability analyzer were used to compare their relative merits and spectral characteristics in the time and frequency domain. Our goal was to establish a benchmark for ourselves as well as others regarding the capabilities of these new frequency control elements. To implement such a comparison, the SAW devices were compared to an HP 10544 bulk oscillator at 10 MHz. Since this type of oscillator generally is several orders of magnitude "quieter" than the SAW oscillators, an HP 8660 frequency synthesizer was used as the reference oscillator for the stability analyzer measurement.

Shown in Fig. 4 is the measured phase noise of three oscillators. The top trace is the single side-band phase noise of a SAW delay line oscillator; the middle trace the noise of a SAW resonator oscillator, and the lower trace the noise of an HP 10544 bulk oscillator multiplied to 300 MHz for comparison. Close to the carrier the bulk oscillator is clearly superior. The SAW delay line is limited primarily by low Q or phase slope, hence, its noise spectrum is the poorest of the three. Similar results have been reported by Parker⁶ and Joseph⁷ in this symposium. The far out phase noise of either the SAW resonator or delay line oscillator can be lower than the multiplied bulk oscillator. This is due to the multiplication of the bulk oscillator's thermal noise floor. As pointed out by Parker, the slope of the close-in phase noise for SAW oscillators can be modeled by an f^{-3} relationship indicative of a rather large flicker noise source.

The same stability information may be obtained by measuring the time domain fractional frequency stability as a function of gate time. Such a comparison for SAW and bulk oscillators is shown in Fig. 5. This type of measurement is in many ways preferred over phase noise. Because the frequency deviation is normalized, any oscillator at any frequency may be compared directly with any other oscillator at a different frequency. Furthermore, relatively inexpensive time domain stability analyzers are available, whereas phase noise measurement equipment is not with the exception of the HP 5390 system.

In Fig. 5 the stability of the SAW and bulk oscillators for short times, i.e., 1 msec, is quite comparable; however, at 1 sec counts the bulk oscillator is 1-2 orders of magnitude more stable. The floor observed is the previously-mentioned flicker effect and is at present the limiting factor in SAW oscillators. To date, the best or lowest flicker floor reported by Adams and Kusters⁵ is 5×10^{-11} . An ambitious goal which appears achievable within the near future is perhaps 1×10^{-11} for a 1 sec count.

Conclusions

The stability of SAW oscillators operating at UHF frequencies has been presented and compared to existing bulk crystal oscillator technology. Our first conclusion is that SAW oscillators are not as stable as bulk

oscillators because of their lower Q and high flicker noise floors. Presently, bulk oscillators are capable of achieving 5×10^{-13} for a 1 sec count, whereas the best SAW oscillator to date has only achieved 5×10^{-11} . In order to improve the stability of SAW oscillators, higher Q's are needed as well as parallel advanced in reducing the high flicker noise of associated oscillator circuitry.

Time domain measurements are favored over frequency domain measurements because (1) oscillator performance may be more easily compared, and (2) time domain measurements are more easily implemented with computing counters. For system applications, however, frequency domain measurements are frequently required.

A time domain measurement system with on-line data processing has been shown to provide excellent results at low Fourier offset frequencies.

Acknowledgements

The authors would like to acknowledge the contribution of data on HP 10544 oscillators by Dr. D. Ricci, and express our appreciation to C. C. Malmuth for her careful composition of this paper.

References

1. E. Hafner, "The Effects of Noise in Oscillators," Proc. IEEE 54, 179 (1966).
2. E. Hafner, "Stability of Crystal Oscillators," Proceedings of the 14th Annual Symposium on Frequency Control, U. S. Army Electronics Command, Ft. Monmouth, New Jersey, pp. 192-199.
3. R. A. Baugh, "Frequency Modulation Analysis with the Hadamard Variance," Proceedings of the 25th Annual Symposium on Frequency Control, U. S. Army Electronics Command, Ft. Monmouth, New Jersey, pp. 222-224.
4. L. Peregrino and D. W. Ricci, "Phase Noise Measurement Using a High Resolution Counter with On-Line Data Processing," Proceedings of the 30th Annual Symposium on Frequency Control, U. S. Army Electronics Command, Ft. Monmouth, New Jersey, pp. 309-317.
5. C. A. Adams and J. A. Kusters, "Deeply Etched SAW Resonators," 31st Annual Symposium on Frequency Control, U. S. Army Electronics Command, Ft. Monmouth, New Jersey, this issue.
6. T. E. Parker, "Current Developments in SAW Oscillator Stability," 31st Annual Symposium on Frequency Control, U. S. Army Electronics Command, Ft. Monmouth, New Jersey, this issue.
7. T. R. Joseph, "SAW Oscillators for Phase Locked Applications," 31st Annual Symposium on Frequency Control, U. S. Army Electronics Command, Ft. Monmouth, New Jersey, this issue.

SAW OSCILLATORS

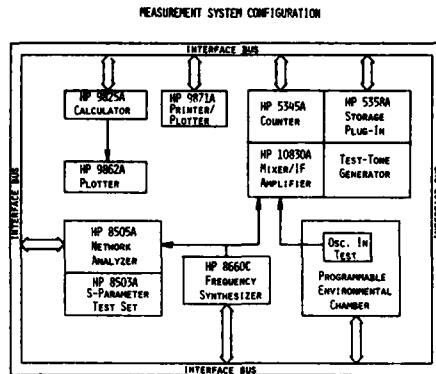


Figure Captions

Fig. 1 Integrated on-line crystal measurements system

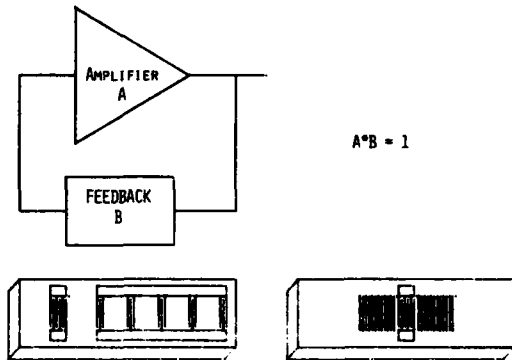


Fig. 2 Typical oscillator circuit and two types of SAW oscillator crystals

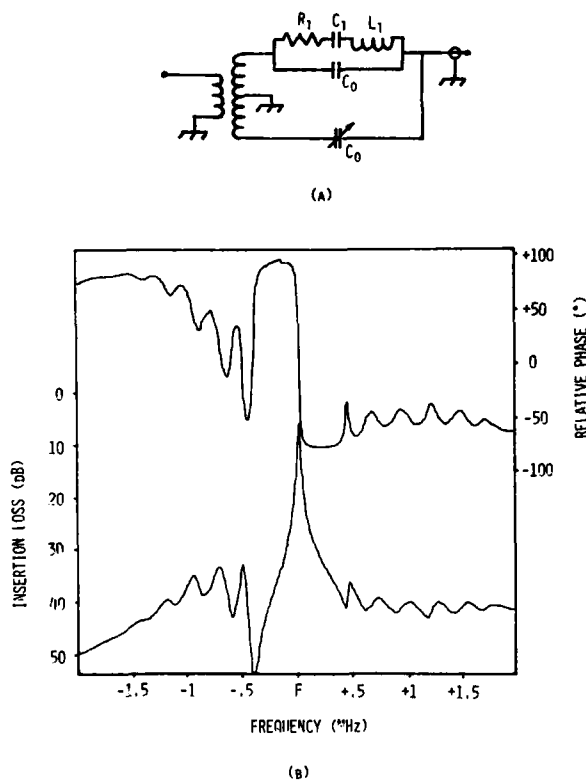


Fig. 3 (a) Feedback network
(b) Measured phase and insertion loss of feedback network

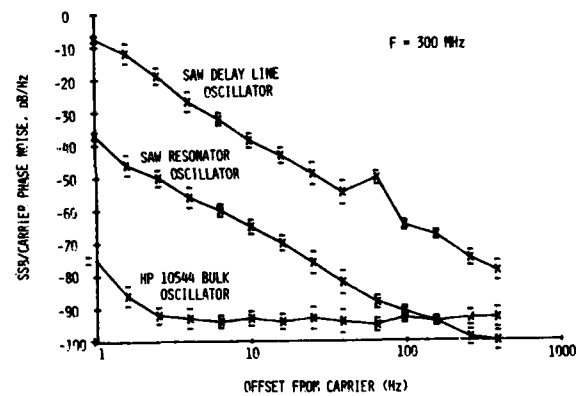


Fig. 4 Single sideband phase noise for SAW oscillator vs HP 10544 bulk crystal oscillator as a function of offset frequency

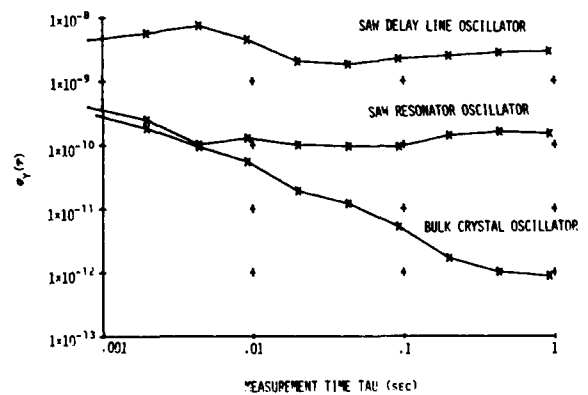


Fig. 5 Fractional frequency stability of SAW oscillators vs HP 10544 bulk crystal oscillator as a function of gate time τ

ULTRA-STABLE LC OSCILLATORS AND THEIR APPLICATIONS IN METROLOGY

Craig T. Van Degrift

National Bureau of Standards
Heat Division
Cryogenic Physics Section
Washington, D.C. 20234

Introduction

It is often stated that, since frequency is the most accurately measurable physical quantity, devices should be developed which convert other physical quantities to frequency. Consequently, there has been widespread effort to develop such devices. This is further stimulated by the need for inherently digital devices for use in automated systems. In this paper, I wish to describe an approach to this problem which is centered on pushing the forty-year old LC oscillator method to its theoretical performance limitations. The advent of the tunnel diode gain element 20 years ago and the relatively new development of high quality microwave capacitors greatly assists in attaining this goal.

Serious use of tunnel diode LC oscillators (TDO's) as transducers originated 12 years ago in the performance of physics experiments at temperatures below 4 K.¹ They were used at frequencies between 1 and 20 MHz and typically had short-term frequency noises of 10^{-7} parts. The traditional problems of LC oscillators - thermal expansion and electrical resistivity changes - are diminished by more than 3 orders of magnitude upon reduction of LC circuit temperature from room temperature to 4 K. If, however, the gain element in the oscillator remains at room temperature, much of this advantage can be lost because of instabilities in the transmission line to the cold LC circuit. Since tunnel diodes work well at low temperatures and in high magnetic fields and also have particularly simple circuits, they are well-suited as gain elements for these applications.

Eight years ago I made some rather obvious engineering improvements² to the existing low temperature tunnel diode oscillator circuits and attained a short and long term frequency noise of about 10^{-9} parts. This, we shall see, is near the theoretical perturbation noise limit of these low-frequency copper LC oscillators for averaging times of a few seconds.

An optimization study of the LC circuit geometry then showed that the best LC circuit for tunnel diode oscillator of a given size is a very low impedance reentrant cavity resonator.³ This type of LC circuit avoids reproducibility problems of conventional LC circuits in room temperature applications, thereby allowing serious use of these oscillators at room temperature. I have recently made room temperature circuits with a frequency noise of 2×10^{-9} parts for 1 sec averaging time. I am now using these reentrant resonator tunnel diode oscillators in a wide variety of applications ranging from temperature measurement using a tiny gas thermometer to rather involved NMR experiment to measure absolute nuclear magnetic susceptibilities.

The remainder of this paper first discusses tunnel diode LC oscillators in general, then reviews the low-temperature 1 - 4 MHz work in greater detail, and finally describes the recent higher frequency

reentrant resonator work and projects currently in progress. Portions of the work reviewed in the next two sections have been previously published in a very detailed paper.²

General TDO considerations

The basic circuit

When properly biased, tunnel diodes⁴ act as non-linear negative ac resistance elements for rf voltages of magnitude less than 30 mV rms. They can be fabricated with characteristic resistance values ranging from -1Ω to $-10 \text{ k}\Omega$ and corresponding characteristic dc bias currents between 50 mA and 5 μA . The gain of this negative ac resistance is then used to overcome the losses existing in a suitably matched LC resonator. The basic circuit is shown in Fig. 1. The only components required besides the tunnel diode and LC resonator are a dc bias resistor, R_2 , a by-pass capacitor, C_B , occasionally a parasitic suppression resistor, R_p , and also occasionally a series resistor R_1 and coupling capacitor C_c . The circuit is powered by a stable dc current flowing through the same coaxial cable used as the rf output. A key element in the design is the maintenance of a high degree of isolation between the circuit and possible external influences and noise. This is accomplished by completely shielding the oscillator and using a large by-pass capacitance to greatly reduce the effect of external rf interference and impedance variations on the oscillator frequency while still permitting a few μV of rf signal to pass out the coaxial cable to the frequency measurement instrumentation.

Frequency Noise

The frequency noise of these oscillators can be divided into the three traditional noise types: addition noise, perturbation noise, and $1/f$ noise. Typically, the addition noise dominates for $\tau < 0.1$ seconds, perturbation noise for τ between 0.1 and 10 seconds and $1/f$ noise for $\tau > 10$ seconds.

The size of the addition noise is determined by a designer compromise between the required output signal amplitude, the maximum allowable circuit power dissipation, and the extent to which the circuit must be isolated from external noise and impedance variations.

Since copper LC₃ resonators have a relatively low Q of the order of 10^3 , the perturbation noise tends to be much greater than for quartz and atomic resonators. On the other hand, the upper limit of driving power is set by completely different physical processes and, in some applications, it may be quite large. Tunnel diodes can provide an rf driving power ranging from as small as 100 pW to as large as 1 mW with 10 nW to 10 μW being most common.

Although I have not yet convincingly observed it, I can easily predict a $1/f$ noise caused by measurable $1/f$ fluctuations in the voltage across the p-n junction of the tunnel diode. These voltage fluctuations can cause fluctuations in the junction capacity and hence in the frequency. For germanium tunnel diodes, the $1/f$ noise floor is roughly at $10^{-9} \nu_{\text{MHz}}/Q$ parts, where ν_{MHz} is the oscillator frequency in MHz. Thus, for low-temperature, low-frequency circuits, $Q \approx 10^3$ and $\nu_{\text{MHz}} \approx 1$, $\sqrt{\langle \sigma^2(2, \tau, \tau) \rangle} \approx 10^{-12}$ while for room temperature reentrant resonator oscillators with $Q \approx 10^3$ and $\nu_{\text{MHz}} \approx 500$, $\sqrt{\langle \sigma^2(2, \tau, \tau) \rangle} \approx 5 \times 10^{-10}$.

I shall point out, however, that since this type of $1/f$ frequency noise is correlated with measurable voltage fluctuations, it is avoidable. One can, in principle, measure the voltage fluctuations and make a correction to the measured frequency.

Low Temperature Low-Frequency Circuits

Circuit for 4 MHz measurements with a capacitor

Since these circuits have already been described in great detail in the literature^{2,3}, I will restrict the discussion here to a summary of their demonstrated performance and a brief description of a digital computer simulation.

Table I summarizes the important performance parameters measured on the circuit shown in Fig. 1. The oscillator geometry allowed the resonator temperature to be varied independently of the remainder of the circuit. As indicated in the table, the resonator temperature coefficient was on the order of 1 ppm/K from 1 K to 15 K. This temperature dependence was caused predominantly by surface impedance changes

occurring in response to changes in the electrical resistivity of the high purity copper wire in the inductor. At these low temperatures, the resistivity of the copper had a minimum at about 10 K and hence the frequency had a maximum at that temperature. For these 4 MHz circuits, the effect of thermal expansion becomes greater than that of surface impedance only at temperatures above 50 K.

The much smaller dependence of the frequency on the "bath" temperature, also given in Table I, reflects the fact that variations in the properties of the auxiliary components of the circuit (R_p , C_B , R_1 , R_2 , C_c and the tunnel diode) are much less important than variations in the primary L or C. An 0.1% K variation in R_2 caused most of the 0.1 ppm/K frequency variation with "bath" temperature.

The frequency dependence on bias current is quite large, as shown in the table, but it is not so large that sufficiently stable current sources are difficult to find. A simple mercury battery with a series current reference resistor monitored by a good voltmeter was adequate to render bias current variations less important than the intrinsic perturbation frequency noise at 40 seconds averaging time. As discussed above, with regard to noise sources, ultimately the $1/f$ fluctuation in diode dc bias should cause difficulties, especially for the higher frequency circuits to be discussed later.

The measured frequency noise of $6 \times 10^{-10}/\sqrt{\tau}$ parts for $1 < \tau < 40$ sec is within a factor of two of the calculated perturbation noise, the discrepancy not being significant compared with the uncertainties of parameters entering into the calculation and with experimental errors caused by measurement system noise of a similar size.

PERFORMANCE CHARACTERISTICS OF A LOW-TEMPERATURE 4 MHz TUNNEL DIODE OSCILLATOR USING A RESONATOR MADE OF COPPER AND SAPPHIRE

Resonator Quality Factor	$Q = 700$	
rf Power Applied to Resonator	$P_{LC} = 0.02 \mu W$	
Total Circuit Power	$P_{\text{total}} = 30 \mu W$	
Dependence on Resonator Temperature	$\frac{\Delta f}{f} / \Delta T_{LC} < 10^{-6} / K$	$\left. \begin{array}{l} 1 K < T_{LC} < 15 K \\ 1 K < T_{\text{bath}} < 4.2 K \\ I = 74 \mu A \\ B < 0.004 T \end{array} \right\}$
Dependence on Bath Temperature	$\frac{\Delta f}{f} / \Delta T_{\text{bath}} < 10^{-7} / K$	
Dependence on Bias Current	$\frac{\Delta f}{f} / \frac{\Delta I}{I} = -6 \times 10^{-6}$	
Dependence on Magnetic Field	$\frac{\Delta f}{f} / B^2 = -3 \times 10^{-4} T^{-2}$	
Frequency Noise	$\sqrt{\langle \sigma^2(2, \tau, \tau) \rangle} = \frac{6 \times 10^{-10}}{\sqrt{\tau}}$	$1 < \tau < 40 \text{ sec}$
Drift	$< 3 \times 10^{-9} \text{ per day}$	$T_{LC} = T_{\text{bath}} = 4.2 K$
Reproducibility of Dependences	better than 10^{-9}	over given range
Thermal Relaxation Time Constant	$< 1 \text{ minute}$	$T = T_{\text{bath}} = 4.2 K$
Warm-up Time to Within 10^{-9} Parts of Final Frequency*	$< 1 \text{ sec}$	$T = T_{\text{bath}} = 4.2 K$

* This depends on thermal linkage of oscillator to its stabilizing temperature bath.

TABLE I

The remaining four entries in the table are upper limits to the quantities shown. Their measurement was limited by factors not related to the oscillator design.

This circuit was used to measure the temperature dependence of the dielectric constant of liquid He at vapor pressure from 0.3 K to 4.2 K with particular emphasis near the superfluid phase transition at 2.1720 K. The capacitor was filled with the liquid He, and a correction, which was quite serious below 0.6 K, was made for a background temperature dependence of the capacitor.

A circuit for 1 MHz measurements with an inductor

A somewhat different low-frequency low-temperature oscillator was then built by David P. Love and myself to measure the surface impedance of metals. A schematic diagram is shown in Fig. 2 in which the equivalent circuit for the tunnel diode is explicitly shown in a dashed box. The 166 turn copper inductor was wound on a 1.27 cm outside diameter x 0.95 cm inside diameter x 10 cm long sapphire tube which had been helically grooved along its outside diameter to accurately position the wire turns.

Resistor R_p was necessary to suppress the parasitic resonance of L_1 and C_1 while capacitor C_2 was necessary to roll-off the high frequency gain of the diode so that helical resonance modes in the coil would not oscillate. In addition, R_2 was necessary to suppress the parasitic resonance of C_1 and the high Q inductance of its superconducting leads. A minor additional difference from Fig. 1 is that a four-element low-pass RC network was necessary to obtain the desired degree of isolation of the oscillator from external influences.

With the entire circuit maintained at a fixed bath temperature, the sample is suspended in the center of the coil from the mixing chamber of a dilution refrigerator. In this manner very small frequency changes associated with the sample (caused by, for example, skin depth changes in a single crystal of copper) could be observed without any correction necessary for the circuit temperature dependence. Figure 3 shows a typical result of such an experiment. The resistance minimum associated with the Kondo effect of 2 ppm Fe impurities in the sample crystal of copper is evident as a frequency maximum at 8.9 K. As shown in the figure, the frequency noise can only be seen when the maximum is enlarged 800 times. The central portion, measured with a counter averaging time of 30 seconds has a noise of 1×10^{-10} parts rms and corresponds to resistivity changes in the copper crystal of only 3×10^{-14} Ω -cm. The applications of this technique as an accurate resistance-to-frequency converter are limited by the great complexity of the temperature dependent anomalous skin effect theory necessary to convert dc resistivity variations to surface impedance changes.

Computer Modeling

Because of the simplicity of these oscillator circuits, accurate modeling of their non-linear behavior on a large-scale digital computer is practical. To demonstrate this, we applied a sophisticated Fortran subroutine which solves simultaneous non-linear differential equations to a simplified version of the circuit in Fig. 2. We assumed that R_2 and L_2 could be neglected and that C_{B1} could be considered an ac short circuit. We also assumed that C_j was connected

in parallel with C_1 and C_{j1} . The resulting circuit had three current loops requiring the numerical solution of three simultaneous differential equations - a linear 2nd order, a linear 1st order, and a non-linear 1st order. The measured tunnel diode I-V curve is used to obtain the non-linear values for $R_n(V)$ and the theoretical voltage variation of the junction capacity is used for $C_j(V)$.

The program determines the oscillation frequency to a part in 10^7 and oscillation amplitude to 0.1% for a given dc bias current and resonator Q. These results for the oscillator of Fig. 2 are shown in Figs. 4 and 5, respectively. The frequency axis in Fig. 4 is scaled in order to indicate the amount (in ppm) by which the actual oscillator frequency deviates from $f_0 = 1/2\pi\sqrt{LC}$. In Fig. 5 is plotted the amplitude of the fundamental Fourier component of the rf current, I_1 , through the diode. This is proportional to the output rf voltage when measured by narrow-band techniques. It can be seen that, at a dc bias current of 350 μ A, an increase of Q above 900 does not produce a greater fundamental signal, but rather generates stronger harmonics. We have seen this effect in the actual circuit. Another effect of the non-linearities is the double-valued nature of these curves near the extremes of bias current. This behavior is associated with a self-rectification of the rf by the diode which causes the actual dc bias voltage across the diode to be dependent on rf amplitude, even for the same circuit bias current. The computer simulation of this effect also quantitatively agrees with our observations of the differences between bias currents at the "drop-out" and "restarting" points.

The accurate simulation of their circuits allow TDOs to be confidently used as a scientific tool. It is very difficult to determine by analytic means alone the extent of validity of various simple analytic formulae, such as $f = 1/2\pi\sqrt{LC}$, often used to describe LC oscillator performance.

High Frequency Measurements Using Reentrant Resonators

Measurements may be made from below 30 MHz to above 2000 MHz using either the capacitive or inductive portions of a low-impedance reentrant cavity resonator. Furthermore, as emphasized earlier, these resonators work well even at room temperature and above. Reentrant resonators with very large capacitance-to-inductance ratios (typically 0.1 to 0.5 farads/henry) work best because their large capacitance greatly diminishes the detrimental effects of the tunnel diode junction capacity. Their high Q (for a given volume and frequency) and simple geometry are, of course, additional advantages.

The circuits used with reentrant resonators can be even simpler than that shown in Fig. 1. If a high quality microwave capacitor is used for the bypass capacitor, resistor K_1 and capacitor C_2 are not necessary to attain adequate isolation at these higher frequencies. Occasionally the parasitic suppression resistor will also be unnecessary. Furthermore, the impedance matching of the diode to the resonator can be performed with a coupling loop or a transmission line transformer. The remaining essential auxiliary circuit components are so small that, if necessary, they can be placed within an SMA connector.

PERFORMANCE CHARACTERISTICS OF A ROOM-TEMPERATURE 90 MHz TUNNEL DIODE OSCILLATOR
USING A COPPER REENTRANT RESONATOR

Resonator Quality Factor	$Q = 600$	} 273 K < T < 300 K I = 1 mA B < 1.3 T
rf Power Applied to Resonator	$P_{LC} = 0.5 \mu W$	
Total Circuit Power	$P_{total} = 50 \mu W$	
Dependence on Temperature	$\frac{\Delta f}{f} / \Delta T = -7.3 \text{ ppm/K}$	
Dependence on Bias Current	$\frac{\Delta f}{f} / \frac{\Delta I}{I} = -8.7 \times 10^{-5}$	
Dependence on Magnetic Field*	$\frac{\Delta f}{f} / B^2 = -1.3 \times 10^{-5} T^{-2}$	} 1 < \tau < 10 sec
Frequency Noise	$\sqrt{\langle \sigma_y^2(2, \tau, \tau) \rangle} = 8 \times 10^{-10}$	
Drift**	$= 2.8 \times 10^{-7} \text{ per day}$	
Reproducibility of Dependences	not yet measured	T = 273 K
Thermal Relaxation Time Constant	$\approx 20 \text{ seconds}$	T = 273 K
Warm-up Time to Within 10^{-9} Parts of Final Frequency	not yet measured	

* Measured at 50 MHz oscillator frequency.

** Still of decreasing magnitude when experiment was terminated.

TABLE II

Table II shows the result of rough measurements of the room temperature performance of an evacuated copper reentrant resonator with an inductor volume of 100 mL. All of these numbers were obtained at 90 MHz except for the magnetic field dependence which was measured at 50 MHz. I expect that further work will quickly reduce the measured long term drift of 2.8×10^{-7} parts/day by about two orders of magnitude, and also that smaller resonators will have much shorter relaxation times. The other entries in the table, however, will probably not be easily improved.

In the following subsections, I will discuss, in rather general terms, applications using these resonators to the measurement of various physical quantities.

Using the capacitive portion of a reentrant resonator

Figure 6 shows a cross-section view of a reentrant resonator and illustrates how it can be used in a tunnel diode oscillator circuit to measure pressure, temperature (as a gas thermometer), acceleration, and linear position changes. The capacitive gap occurs between the top of a central post and the bottom of a thick diaphragm. The single-turn inductor is formed by an annular wall connecting the diaphragm to the central post so that the rf magnetic field lines therefor encircle the post. The capacitive gap is very small. It is typically set between 1 and 10 μm . The remaining circuit components are placed within a hollow region inside the central post and drive a coupling ribbon which emerges from the central post and is attached to the outer wall.

As shown in Fig. 6, pressure is sensed if a gas is allowed to push on the top side of the diaphragm. Alternatively, if gas is trapped in a closed volume with the diaphragm as one wall, a gas thermometer is

formed. Mass loading of the diaphragm can make it sensitive to acceleration while mechanical linkage to it allows very precise measurement of linear motion. Dielectric constants can be measured if the upper capacitor plate is made rigid and a sample is placed in the capacitor gap.

We have made a small (1.9 cm dia. x 5.1 cm long) secondary gas thermometer designed to work best in the temperature range of 10 to 30 K where conventional germanium and platinum resistance thermometers have poor sensitivity. An illustration of the dependence of its frequency on temperature is given in Fig. 7. The measured oscillator temperature dependence is given by the solid line while the gas pressure contribution and the combined contributions of gas pressure and thermal expansion are given by the dashed line and by the dash-dot line, respectively. The difference between this last curve and the observed behavior is the right sign and magnitude to be explained by the temperature dependent surface impedance of the copper in the resonator walls. This thermometer has already demonstrated a resolution of 100 μK at low temperatures and 50 μK at room temperature. We are presently making an improved version which will also contain a pressure sensing oscillator and will dissipate less than 15 μW of power. It remains to be seen how stable these can be made with respect to ageing and thermal and mechanical shock. It might be necessary to also install an auxiliary oscillator which senses the surface impedance changes of superconductive temperature fixed points between 1 and 20 Kelvin. These would then provide internal calibration points.

We have also placed a copper reentrant resonator TDO built with a rigid upper capacitor plate in an accurate humidity generator to determine its potential as a hygrometer. The results were encouraging but it is clear that the capacitor surfaces must be plated with an inert material to make them chemically stable in the presence of moist air.

Using the inductive portion of a reentrant resonator

The manner in which the inductor of a reentrant resonator can be used to do absolute nuclear magnetic resonance (NMR) is shown in Fig. 8. By absolute NMR, I mean accurate measurement of the absolute magnitude of resonant nuclear susceptibility. Errors in sample filling factor, resonator Q, and overall system transfer function prevent the accurate measurement of nuclear susceptibilities by ordinary techniques. A technique developed at NBS by Donald B. Utton and myself uses the calculable reentrant resonator geometry to permit determination of the sample filling factor with an accuracy of 100 ppm. In addition, our method uses measurements of the frequency pulling, rather than the amplitude change, of a very stable 50 MHz tunnel diode oscillator to avoid the other errors. This method is only practical because of the exceedingly high frequency stability of our tunnel diode oscillator.

The oscillator circuit in Fig. 8 is coupled directly to the capacitor rather than through a loop in the inductor. This is done in order to preserve the calculable nature of the inductor. If less than 100% coupling is desired, a transmission line impedance transformer can be used between the diode and reentrant resonator capacitor.

An evaporated thin film insulator can be placed in the outer edge of the capacitor gap to form a vacuum seal between the capacitor surfaces. This then greatly reduces the sensitivity of the oscillator to the dielectric properties of the sample and also to vibrations and temperature gradients.

We are now exploring the full capabilities and limitations of this method of NMR detection. In addition, we are considering the straightforward uses of a tunnel diode oscillator with a reentrant resonator inductor to measure surface impedance and non-resonant paramagnetic susceptibility.

Difficulties

I shall temper my enthusiasm for these LC oscillators by pointing out three problem areas with which I see difficulties:

- 1) vibration sensitivity
- 2) sensitivity to thermal gradients
- 3) sensitivity to submonolayer desorption and adsorption on the capacitor surfaces.

In some applications the first two of these problems may be greatly reduced by placing an insulating spacer in the capacitor gap. One thereby, however, compromises the principal of single material construction of the reentrant resonator - hence jeopardizing the inherent reproducibility of the design. The third problem is similar to the packaging problems of quartz crystal resonators and can probably be overcome using the same packaging techniques.

Devices presently being built and tested at NBS

There are five projects currently in progress to determine the extent to which the ideas presented in this paper can be realized in practice. They are:

- 1) Construction of an improved low-temperature secondary gas thermometer which also indicates either ambient pressure or gives a readout of the transition temperature of superconducting fixed points. It will operate at 500 MHz and be 2.2 cm dia. and 6 cm long.

- 2) Construction of a room-temperature thermal expansion thermometer and pressure gauge. It is to operate at a frequency of 1 GHz and to have a size of 1 cm dia. x 3 cm long.

- 3) An NMR susceptibility measurement with 100 ppm accuracy at 273 K. This will be done at frequencies between 20 and 50 MHz and will have a 100 mL sample volume.

- 4) A measurement of NMR susceptibility and pressure vs. temperature for solid ^3He between 20 mK and 100 mK. This will be done at 150 MHz and will use only a 1 mL sample volume.

- 5) A long term stability test of a large (100 mL inductor volume), carefully cleaned and annealed room temperature 100 MHz oscillator.

We will probably soon also start experimenting with tunnel diode oscillators which use superconducting reentrant resonators as well as ones which use the "electrodeless" quartz resonators of Prof. R. Besson.

Conclusion

I have shown how tunnel diode oscillators using non-superconducting LC resonators can be made which are limited by fundamental noise processes at the 10^{-9} to 10^{-10} parts level. They operate from below 1 MHz to over 500 MHz, from below 0.3 K to over 300 K, and in magnetic fields from 0 to at least 1.3 tesla. They have circuits which are so simple that a detailed digital computer modeling of their performance vs. dc bias, temperature, and magnetic field is practical. Finally, I have shown how they are naturally suited for the task of converting changes in a wide variety of physical quantities into frequency changes.

References

1. C. Boghosian, H. Meyer, and J. E. Rives, Phys. Rev. **146**, 110 (1966).
2. Craig T. Van Degrift, Rev. Sci. Instrum. **46**, 599 (1975).
3. Craig T. Van Degrift, Rev. Sci. Instrum. **45**, 1171 (1974).
4. Higher impedance versions are sold as "back diodes."
5. Craig T. Van Degrift and David P. Love, Low Temperature Physics - LT-14, Vol. 4, Eds. Matti Krusius and Matti Vuorio (North-Holland Press, Amsterdam, 1975), p. 309.
6. Craig T. Van Degrift, Ph.D. Thesis (University of California, Irvine, 1974), Available from University Microfilms, 300 North Zeeb Road, Ann Arbor, MI 48106. Order No. 74-19243.
7. David P. Love, C. T. Van Degrift, and W. H. Parker, Low Temperature Physics LT-14, Vol. 3, Eds. Matti Krusius and Matti Vuorio (North-Holland Press, Amsterdam, 1975), p. 106.
8. Craig T. Van Degrift and Donald B. Utton, "Magnetic Resonance Detection Method and Apparatus", Application for U.S. patent Serial No. 774,094, March 3, 1977. We will soon publish a detailed description of this work.
9. Raymond Besson, "A New 'Electrodeless' Resonator Design", These proceedings, (1977).

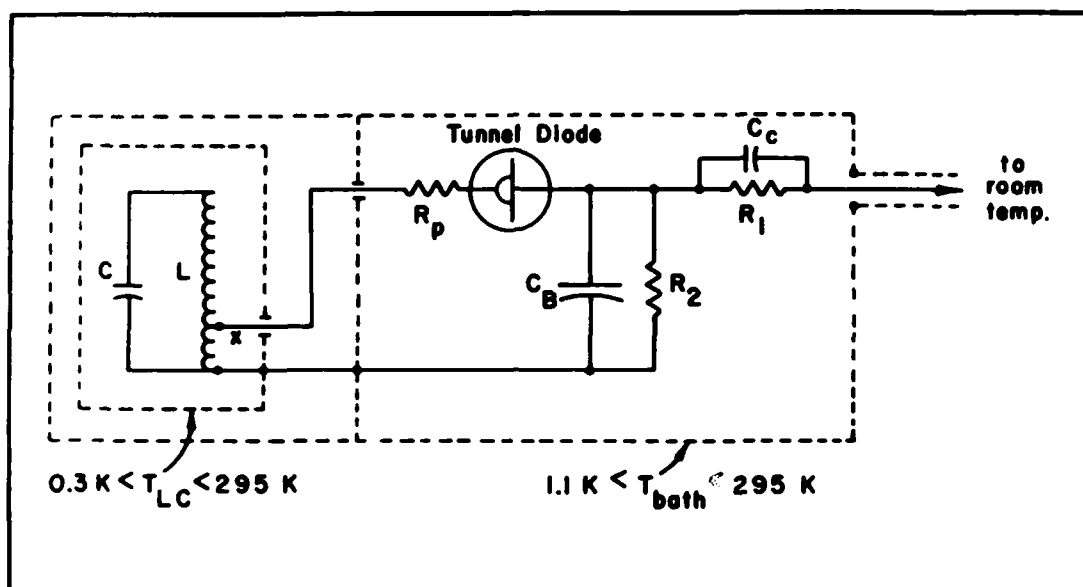


Fig. 1 4 MHz circuit used at low temperatures to study the dielectric constant of liquid ^4He . The performance of this circuit is summarized in Table I in the text. The circuit was constructed with the LC resonator thermally isolated from other components as is schematically illustrated in this figure.

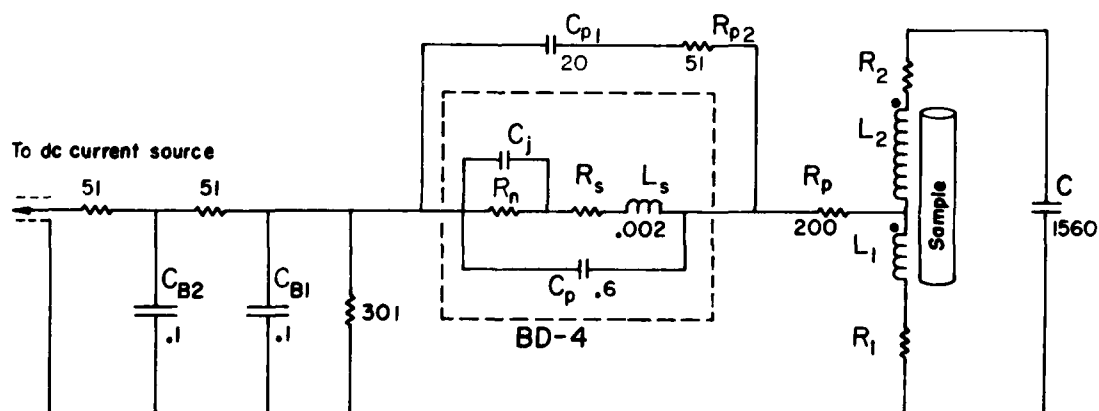


Fig. 2 700 kHz circuit used at low temperatures to study the inductive reactance of samples suspended within its hollow coil form.

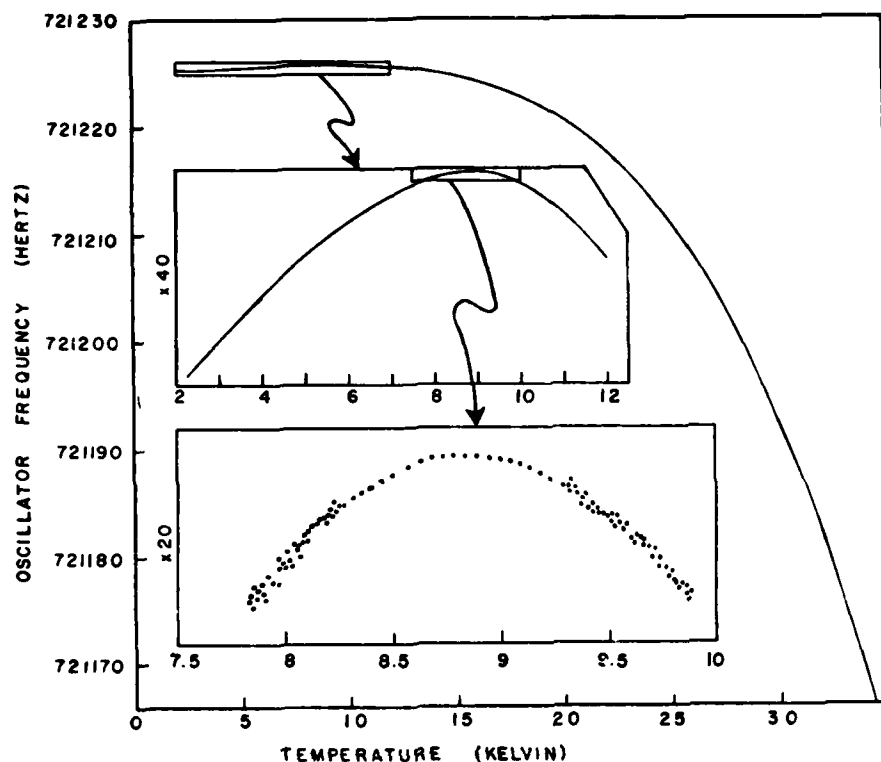


Fig. 3 Variation of the frequency of the oscillator in Fig. 2 with sample temperature while maintaining the circuit temperature constant. The sample is a single-crystal rod of high-purity copper with the 100 direction aligned along the coil axis. This plot qualitatively mirrors the dc resistivity variation of the copper with temperature-increasing resistivity causes greater surface inductance and hence lower frequency. The maximum at 8.9 K corresponds to the well-known Kondo minimum in dc resistivity which results from 2 ppm of iron impurity in the copper. The data in the region of minimum have been plotted with increasing resolution in the two inserts. Upon an 800-fold enlargement, the frequency noise of 2×10^{-10} parts can be seen. The counter averaging time, which was normally 3 seconds, was increased to 30 seconds during the pass through the minimum.

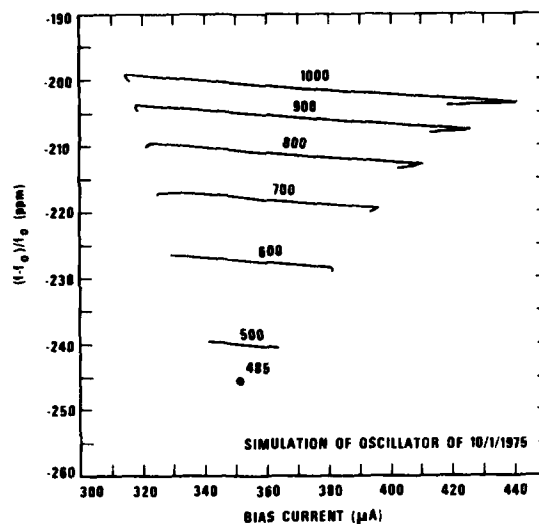


Fig. 4 Calculated deviation of the oscillator frequency from $f_0 \equiv 1/2\pi\sqrt{LC}$ (for the circuit shown in Fig. 2) plotted vs. the total circuit dc bias current for different values of resonator quality factor. The oscillation threshold is at $Q = 485$ and a bias current of $350 \mu A$. Where these curves are double-valued, the upper portions give the steady-state behavior while the bias current values at the end points of the lower portions give the self-starting region of the oscillator at each Q value.

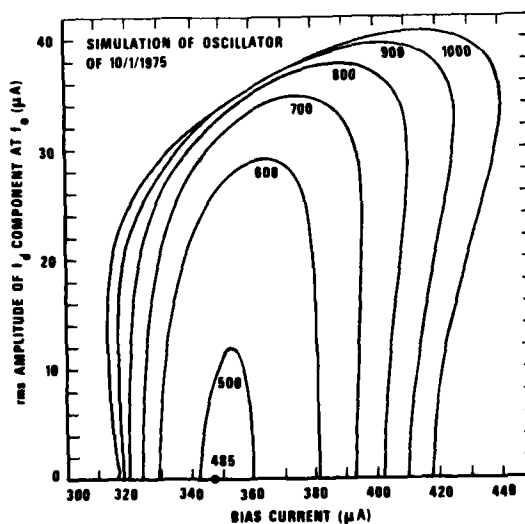


Fig. 5 Calculated rms amplitude of the fundamental Fourier component of the rf current in the diode (for the circuit shown in Fig. 2) plotted vs. the total circuit dc bias current for different values of resonator quality factor. For Q -values which are twice as great as the threshold Q of 485 the limiting effect of the diode non-linearities is particularly evident. The interpretation of these curves where they are double-valued is the same as given in the caption of Fig. 4.

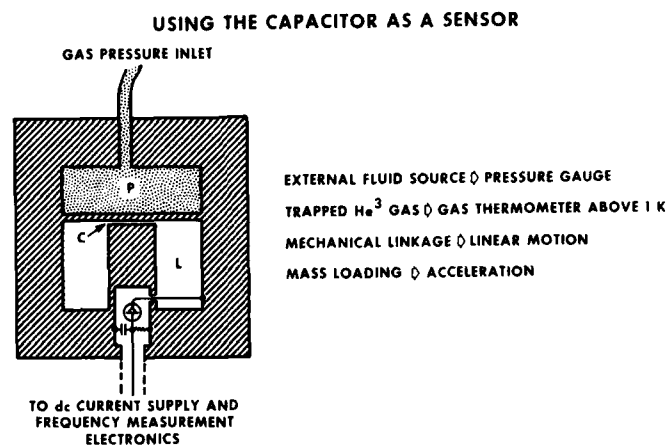


Fig. 6 An illustration of how the capacitive region of a reentrant resonator can be used in transducer applications.

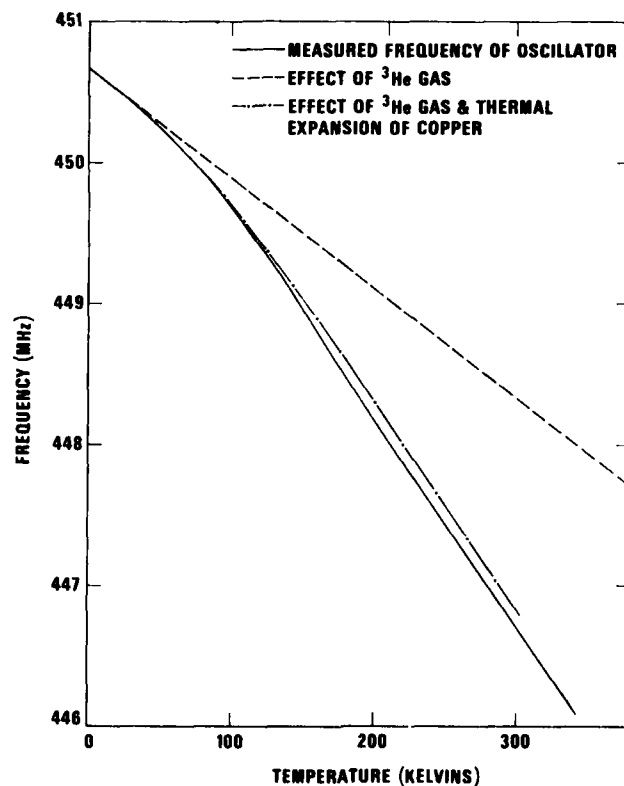
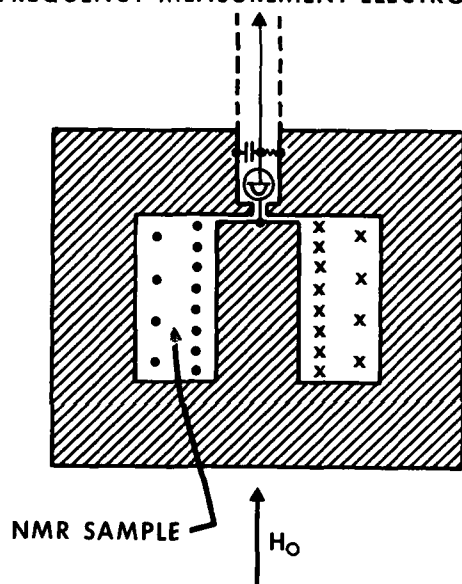


Fig. 7 The measured performance of our first prototype low-temperature gas thermometer. Although the total frequency change was only 1% from 1 K to 315 K, the frequency noise of only 2×10^{-9} parts permitted a temperature resolution of between 50 and 100 μ K over the whole temperature range. The effects of gas pressure alone and of gas pressure plus resonator thermal expansion are also shown.

USING THE INDUCTOR AS A SENSOR

TO dc CURRENT SUPPLY AND
FREQUENCY MEASUREMENT ELECTRONICS



- USE OF THE FREQUENCY PULLING OF A MARGINAL OSCILLATOR INSTEAD OF ITS rf AMPLITUDE CHANGE ELIMINATES SYSTEMATIC ERRORS CAUSED BY AMPLIFIER AND DETECTOR CALIBRATION DRIFT.
- USE OF THE REENTRANT RESONATOR WITH ITS CALCULABLE INDUCTOR GREATLY REDUCES ERRORS IN THE SAMPLE FILLING FACTOR.

Fig. 8 An illustration of how the inductive region of a reentrant resonator can be used to measure absolute nuclear magnetic resonance susceptibilities.

FREQUENCY SYNTHESIZERS FOR AIRBORNE RADARS

F. W. Hopwood
J. P. Muhlbaier
H. Rossman
Westinghouse Electric Corporation
Baltimore, Maryland

Summary

The performance of coherent, multi-mode airborne radars has become dependent on microwave frequency synthesizers for the generation of transmit and local oscillator signals.

Various synthesizer techniques which might be appropriate to the radar application are discussed, along with a description of the Westinghouse "Building Block" synthesizers.

Phase noise is a driving consideration when selecting the synthesis technique. The phase noise requirements associated with the coherent mapping and downlook air-air radar modes are discussed. Simple relationships between radar performance and synthesizer parameters are presented.

Certain radar modes require very rapid frequency change and phase settling, over large frequency excursions. Since phase lock loops are required for noise and spurious reasons, new phase lock circuit techniques are required to achieve the switching speed. A digital technique for switching a phase locked synthesizer over 80 MHz in ten microseconds is described, and data is presented.

The Westinghouse synthesizers are manufactured on polyimide glass which is a high temperature copper clad material combining many of the best properties of thick film and printed circuit technologies. It allows the use of chip, discrete, and packages and printed components on the same assembly. Mechanical and electrical properties of the material are discussed, as well as the packaging methods.

Introduction

The performance of multi-mode airborne radars has become increasingly dependent on the use of microwave synthesizers to generate the transmitted frequencies. The advantages of the use of multiple frequencies include reduced susceptibility to mutual interference, as well as improved detection capability when used in such applications as missile threat detection, ground mapping, and terrain follow/avoidance.

The requirements and mechanizations appropriate to the radar synthesizers differ in some important respects from the general purpose synthesizers seen in test equipment and communications systems. The most apparent difference is the required channel density. It is unusual for a frequency change of less than a reciprocal pulse width to affect radar performance with respect to detection capability or mutual interference. Since one microsecond is a typical pulse width, we find that the radar synthesizer resolution is often five to ten MHz. General purpose microwave synthesizers usually have resolutions of about one kHz. Radar fractional bandwidths are typically 2 to 10 percent, while general purpose synthesizers might cover a decade.

An important radar parameter is frequency switching speed. Frequency switching over hundreds of MHz in ten microseconds is typically required. Since phase

lock loops are necessary to meet noise and spurious requirements, the switching speed makes necessary new and special phase lock circuitry.

Inclusion of a downlook air-air mode requires a phase noise floor of -120 to -125 dB/Hz at X band. This is lower than most commercial synthesizers reach, but is similar to some communications needs. Spurious (-80 dB) and low frequency stability requirements are comparable to those of general purpose synthesizers.

Stability Requirements

Raven¹ and others have described the requirements for radar oscillator stability. Our intent is to relate them to quantities which are familiar to the RF hardware designer, and to point out those quantities which most effect hardware decisions. We place particular emphasis on the stability problem since it, more than any other consideration, drives the choice of synthesis technique. It is important to note that there are many synthesizer types that meet all of our requirements except phase noise.

The stability requirements are of simple origin. In the downlook air-air mode, we wish to detect targets whose doppler frequencies differ from the large "clutter" returns from the earth. Our stability requirement arises from the fact that we wish to observe only small targets or receiver noise when the clutter returns are at their maximum power level P_{CL} . The STALO phase noise floor produces noise sidebands on the clutter returns observed in the receiver. These must be small compared to receiver noise.

This has the effect of imposing a maximum effective multiplication ratio on the STALO, assuming that our phase noise is that of a crystal oscillator and frequency multiplier. If $\mathcal{L}_x(f)$ is the crystal oscillator white phase noise floor, $K_{TO}F$ is the receiver noise, N is the multiplication ratio, and P_{CL} is the maximum clutter level at the receiver front end, the STALO stability criterion is

$$N^2 \mathcal{L}_x(f) < \frac{K_{TO}F}{P_{CL}} \quad (1)$$

For example, if $\mathcal{L}_x(f)$ is -165 dB/Hz, noise figure = 2 dB, P_{CL} = -50 dBm, the STALO noise floor must be -122 dB/Hz at X band and N must be less than 141.

The air-air mode also imposes a Q or bandwidth restriction on the oscillator. The oscillator half bandwidth must be small enough so that the phase noise floor is reached at modulation rates which are compatible with the spectral width of the clutter returns. These are a function of aircraft velocity, antenna characteristics, geometry, and other considerations. Typical X band air-air radars require half-bandwidths of 2000 Hz, which is not severe for VHF crystal oscillators.

¹R. S. Raven, "Requirements on Master Oscillators for Coherent Radar", Proceedings of IEEE, Vol. 54, No. 2, Feb. 1966.

The "Doppler Beam Sharpening" mode, used for air-ground mapping, imposes a low frequency stability requirement. We arrive at crystal Q requirements with a few simple considerations. This radar mode achieves azimuth resolution by observing the slowly changing doppler returns from off-boresight targets with a digital correlator and filter band. Observations are made over an integration time T_i . The matched filter bandwidth is $1/T_i$. If the system dynamic range is D, noise sidebands observed in the receiver must be at least (D) below the carrier in filter bandwidth $1/T_i$, for rates greater than half the filter bandwidth. (See figure 1) The phase noise requirement for independent noise sources is then:

$$\mathcal{L}_i(f) \left(f > \frac{1}{2T_i} \right) < \frac{T_i}{D} \quad (2)$$

Integration time T_i is function of resolution, aircraft velocity, and geometry, and may vary significantly. Allowing T_i to assume all values less than T_{iMAX} results in:

$$\mathcal{L}_i(f) \left(f > \frac{1}{2T_{iMAX}} \right) < \frac{1}{2fD} \quad (3)$$

Low frequency instabilities $\mathcal{L}(f)$ appearing on both transmit and LO receive significant cancellation as observed in the receiver. The oscillator stability requirement is relaxed by the amount of this cancellation. The ratio of common and independent stability requirements is:

$$\frac{\mathcal{L}_i(f)}{\mathcal{L}_c(f)} = 4 \sin^2(\pi f T_{RT}) \quad (4)$$

T_{RT} is the radar round trip time. The common phase noise requirement is then approximately (figure 2):

$$\mathcal{L}_c(f) < \frac{1}{8\pi^2 D T_{RT}^2 f^3} \quad (5)$$

for

$$\frac{1}{2T_{iMAX}} < f < \frac{1}{\pi T_{RT}}$$

and

$$\mathcal{L}_c(f) < \frac{1}{8Df} \quad (6)$$

for

$$f > \frac{1}{\pi T_{RT}}$$

The modulation frequencies of interest generally lie in the $1/f^3$ region of crystal oscillator-multiplier sources, whose noise in that region is:

$$\mathcal{L}_x(f) = \frac{F_x^2 f_1}{4Q^2 f^3} \mathcal{L}_{sc}(f_1) \quad (7)$$

Here, F_x is the transmitted frequency, Q is the crystal oscillator loaded Q, $\mathcal{L}_{sc}(f_1)$ is the $1/f$ phase noise of the crystal oscillator sustaining circuit observed at modulation rate f_1 . Equating (5) and (7) results in the crystal oscillator Q requirement:

$$Q > 2\pi F_x T_{RT} \sqrt{\frac{D f_1 \mathcal{L}_{sc}(f_1)}{2}} \quad (8)$$

Note that the Q is a function of round trip time and dynamic range but is independent of map resolution. A typical X band airborne mapping radar has 250

microsecond round trip time and 30 dB dynamic range. Typical sustaining circuit $1/f$ noise is -120 dB/Hz at one Hz. This results in a loaded Q requirement of 350, which is certainly modest for crystal oscillator.

We have assumed the frequencies of interest to be above those dominated by oscillator thermal effects. Most mapping radars have integration times of less than ten seconds. The lowest sideband frequency of interest is then 0.05 Hz, which is generally within $1/f^3$ region for oscillators with modest temperature control.

It is convenient to measure the low frequency stability in the time domain, using computing counter techniques. Using the frequency - time transforms in NBS Tech Note 632², the time domain stability requirement for common oscillators becomes:

$$\sigma_y^2(\tau) < \frac{\ln 2}{32\pi^2 F_x^2 T_{RT}^2 D} T_{iMIN} < \tau < T_{iMAX} \quad (9)$$

For example, an X band coherent mapping radar with 30 dB dynamic range and 250 microsecond round trip time requires low frequency stability of 5.9×10^{-10} over observation times of perhaps 0.1 to 10 seconds (figure 3).

Synthesizer Building Blocks

Westinghouse has developed a "Building Block" approach to the synthesizer problem. Several synthesizers have been constructed recently, each from similar hardware, but having a diversity of performance parameters including operating band, channel density, bandwidth, and switching speed. The use of common "Building Blocks" such as crystal oscillators, frequency multipliers, and VCO's offers significant cost and time savings, particularly for small programs or emerging, highly competitively programs where new and extensive hardware development can have severe impact.

Airborne Radar synthesizers must need a diversity of requirements to serve the various radar system modes. Some typical requirements which the "Building Blocks" are designed to accommodate are:

- X band transmit frequency
- 300 to 600 MHz coverage
- Phase noise floor -120 to -130 dB/Hz
- Short term stability: 5×10^{-10} (0.1 to 10 sec)
- 5 to 50 MHz channel spacing
- Frequency switching: 10 microseconds
- Spurious: -50 to -90 dB

There are a large number of synthesis techniques to be considered for the radar application. An excellent survey of the subject has been conducted by Manassewitsch.³ While no one of the schemes he presents solves the radar problem, we find that much of his descriptive material is useful and his analyses of phase noise and spurious are appropriate for determining whether a particular synthesis technique will meet the radar requirements. His analyses and measurements verify the fact that many of the commonly used synthesis techniques fail to meet the radar phase noise and band cover requirements.

²J. H. Shoaf, D. Halford, A. S. Risely; "High Frequency Specification and Measurement: High Frequency and Microwave Signals", NBS Tech Note 632, January 1973, pp. 33-46.

³Manassewitsch, V., "Frequency Synthesizers Theory and Design", John Wiley, 1976.

Figure 4 is an example of the frequency synthesizers have been constructed from an assortment of basic "Building Blocks". It is very simple in concept and provides excellent performance in the airborne radar application. The mechanization provides 18 channels spaced 20 MHz over 360 MHz at X band. The phase noise floor is essentially that of the crystal oscillator-multiplier. A noise floor of about -124 dB/Hz is readily achieved.

The VCO, typically a Gunn device, provides the receiver first LO signal at frequencies $8640 \pm 10 \alpha$, where α takes on values (1, 3, ..., 17). The VCO is mixed with the multiplier output (8640 MHz) to produce phase lock IF frequencies ranging from 10 to 170 MHz. These, in turn, are mixed with 0 (dc current), 60, or 120 MHz, then divided by 1, 3, 5 as appropriate. The resultant 10 MHz second IF is mixed to I-Q video to develop phase lock error signals.

The transmit frequency is developed by simply mixing the VCO frequency with 660 MHz, which is also the receiver first IF frequency. The receiver second LO (720 MHz) and third LO (60 MHz), as well as various system clocks, are generated by the multiplier-divider chain. High speed frequency switching (less than 10 microseconds) is achieved by the use of a simple digital control mechanism, to be described later. Various peripheral features, such as passive or active chirp, clutter positioning, beacon offset, can be conveniently added as required by the specific radar application.

This synthesis technique, while not particularly unusual, has the advantages of offering a fairly large number of channels (18) while being simple and quite inexpensive. It has one significant limitation in that it cannot readily provide very close frequency separation without forcing the phase lock second IF (10 MHz in the example) to be too low in frequency. An IF of much less than 10 MHz reduces the available loop bandwidth, and results in slow acquisition time and difficulty in suppressing the VCO phase noise. IF's of 10 MHz and above result in 20 MHz channel spacing, and phase noise and transient response which are satisfactory for many radar applications.

The "Iterative Mix and Divide" synthesizer described in figure 5 finds application in systems which require channel spacing less than about twenty MHz. This mechanization consists of a cascade of N identical synthesis sections, resulting (4^N) channels. Each section mixes its input at approximately F_0 with either of $3F_0$, $3F_0 + \Delta F$, $3F_0 + 2\Delta F$, $3F_0 + 3\Delta F$. The sum of these is filtered and divided by four, then passed to the next stage. The final output can be at either F_0 or $4F_0$. If the output is at F_0 , the channels range from F_0 to $F_0 + \Delta F$. Adding additional stages serves to pack the output bandwidth more tightly, but does not change the bandwidth coverage.

This synthesis technique is particularly attractive for radar application for several reasons. The final frequency divider limits the phase noise floor to about -150 dB/Hz if commercial counters are used. Healey¹ reports that this can be significantly improved with a simple "cleanup" technique. The phase noise floor is essentially that of the last counter, regardless of the number of cascaded sections, since additional stages phase noise is divided by 4, 16 etc. The phase noise floor of a large number of stages is only about 0.28 dB greater than that of one stage if they are equal and uncorrelated. The more important limitation is that of available bandwidth, which is

limited to about 10 or 15 percent if reasonable spurious are required. This limits the available output bandwidth to about 100 to 150 MHz if commercially available counters are used. While the Iterative Mix and Divide synthesizer has a number of advantages, its bandwidth limitation makes it appropriate for only a limited number of radar applications.

Those radars which require a large number of channels spread over a large bandwidth can use a combination of synthesis techniques. Figure 6 describes a hybrid synthesizer which uses a simple mix and divide synthesizer to obtain a narrow (5 MHz) frequency steps, with frequency offsets to achieve large bandwidth (800 MHz). The mix and divide synthesis develops any of 16 frequencies covering 80 MHz bandwidth near 450 MHz. The VCO at about 8 GHz is converted to VHF, then compared with one of the sixteen mix and divide frequencies. The difference is converted to 40 MHz using either dc, 160, or 320 MHz. Using this relatively simple hardware, we obtain 160 channels spaced 5 MHz over 800 MHz at X band. The phase noise is essentially that of the crystal oscillator-multiplier. A noise floor of about -124 dB/Hz is typical.

Frequency Agility

A common radar synthesizer requirement is for very rapid switching through psuedo-random patterns of transmit frequencies. The patterns may cover hundreds of MHz with switching times of five or ten microseconds to five degree phase error. It is difficult to reliably switch conventional phase lock circuitry at the required rate due to dwell time limitations. The dwell time problem is described by Gardner² and others. A conventional means for rapid frequency change is to sweep the VCO to the correct frequency, phase lock, and use a quadrature lock indicator to disable the sweeper. The sweep rate must be such that the VCO dwells within a loop bandwidth of the correct frequency for several (perhaps ten) reciprocal bandwidths. If the VCO searches linearly over bandwidth B_s for time T_s , to find loop bandwidth B_L , reliable lockup is achieved when

$$\frac{B_s}{T_s} < \frac{B_L}{10/B_L} \quad (1)$$

or

$$T_s > \frac{10B_s}{B_L^2} \quad (2)$$

For example, to search a 600 MHz tuning bandwidth using a 1 MHz servo requires about 6 milliseconds. This compares unfavorably with our ten microsecond requirement. One can reduce the search time by prepositioning and restricting the sweep range. This, however, is done at the expense of temperature sensitivity and possibly the need for periodic "pot" adjustments which are generally unacceptable for military applications.

There are an array of analog techniques which might be used to hasten the frequency change. Various discriminators, filters, detectors, etc. have been proposed and investigated in some detail, and show some promise.

A more attractive approach has been to use one of a number of digital control schemes, or possibly a combination of digital and analog hardware. An important consideration is whether or not to use a digital-to-analog converter (DAC), since one of the required resolution (perhaps 12 bits) might be prohibitively expensive in some applications. The use of a

¹D. J. Healey, personal correspondence

²Gardner, F.M., "Phaselock Techniques", John Wiley, 1966.

suitable DAC and control logic can, however, yield excellent performance.

A very simple fast acquisition scheme is described in figure 7. It utilizes a fast sweep on the VCO, with a conventional zero-frequency discriminator (ZFD) to disable the sweep as the VCO is driven to the correct frequency. The final phase lock employs I-Q mixers to develop quadrature video. This drives logic level comparators and the "C" and "D" inputs of a "D" flip-flop. This logic element functions by transferring the "D" input state to the output whenever the "C" input is strobed. When the phase lock IF frequency W_{IF} is above reference frequency W_R , the output is always "1". W_{IF} less than W_R results in a "0". Thus, the flip-flop changes state as the VCO is swept through W_R . This disables the VCO and allows very rapid phase lock. Waveforms (figure 8) associated with such a scheme show that the ten microsecond switching time is readily achieved. Note that this type of mechanization avoids the need for alignment, does not require linear VCO tuning, and is quite insensitive to temperature changes. These features make it particularly suitable for military applications.

Packaging Considerations

The Westinghouse synthesizers are packaged using polyimide substrates on aluminum web heat sinks. The board material is similar to standard printed circuit boards except that polyimide resin is used to bond the layers of glass cloth. The result is a very high temperature bond which does not melt at normal solder temperatures as does epoxy. Chip resistors, chip capacitors, and other small devices can therefore be used, resulting in substantial size reduction. The material is somewhat lossy, averaging 1.6 dB/wavelength from 100 MHz through X Band. While higher than desirable, the loss is acceptable in many applications through about 2 GHz since it is much less expensive to use than alternates such as ceramic. Thermal conductivity of the polyimide material is similar to that of epoxy glass boards, and is much less than ceramic. Consequently, appropriate care must be used when using devices with significant power dissipation. Dielectric constant (4.8) is similar to that of epoxy glass. Figure 9 is an example of synthesizer construction on polyimide substrates.

Conclusion

A family of frequency synthesizers for airborne radars has been configured. The use of common "Building Blocks" for various applications allows rapid and cost effective response to quick reaction or low quantity applications. Rapid frequency switching is accomplished by appropriate combination of digital and analog techniques. The use of polyimide substrates is cost effective, although its relatively high loss may make it inappropriate for some applications.

Stability Criterion for Mapping Radar

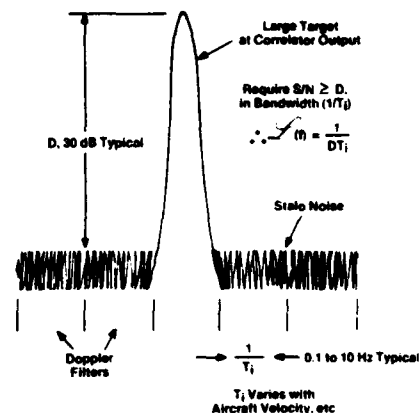


Figure 1

Frequency Domain Stability Requirement for Airborne Mapping Radar

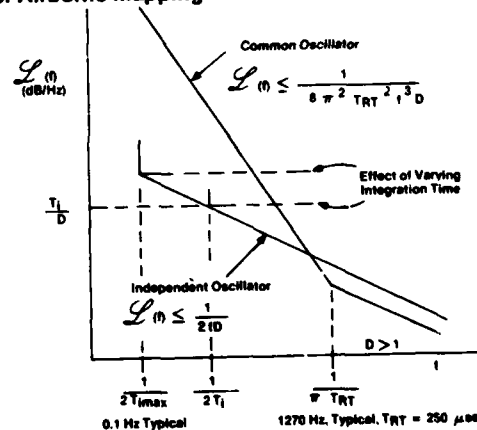


Figure 2

Time Domain Stability Requirement for Airborne Mapping Radar

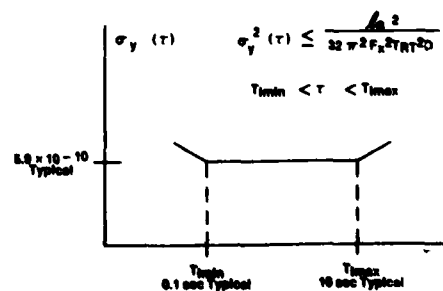


Figure 3

18-Channel Radar Synthesizer Covering 360 MHz at X-Band

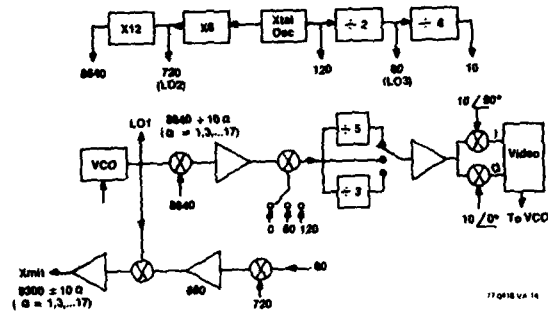


Figure 4

Logic Waveforms for Digital Control

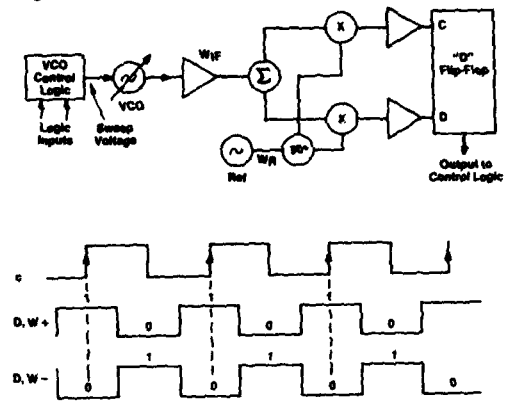


Figure 7

Iterative Mix and Divide Synthesis

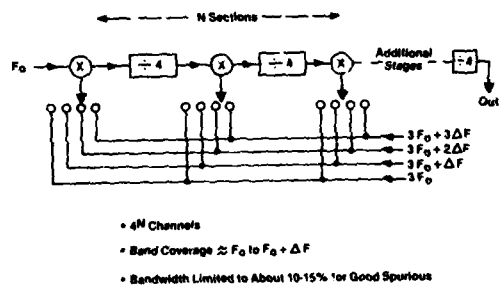


Figure 5

Switching Time

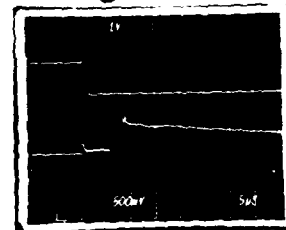


Figure 8

77-0418-VA-18

160-Channel Radar Synthesizer Covering 800 MHz at X-Band

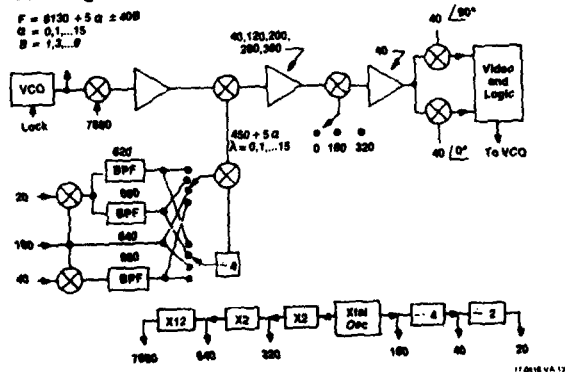


Figure 6

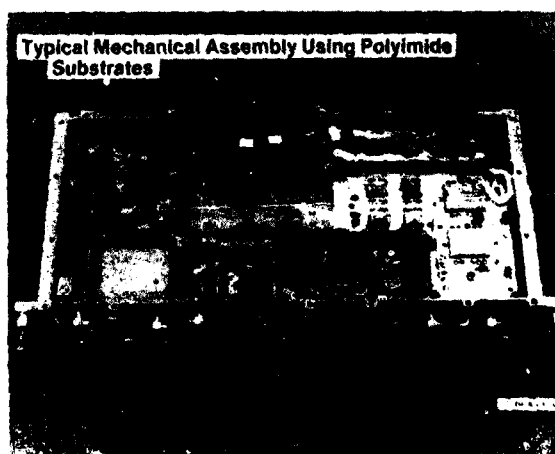


Figure 9

FREQUENCY TOLERANCE LIMITATIONS WITH LOGIC GATE CLOCK OSCILLATORS

John D. Holmbeck
Northern Engineering Laboratories, Inc.
Burlington, Wisconsin

Summary

Some expensive design errors are arising on a large scale due to lack of fundamental understanding of resonant circuit theory and its place in crystal oscillator design by many designers of digital integrated circuits for clock oscillators. In some cases they are adequate for their intended purpose but the published data sheets do not point out the limitations. With better understanding of oscillator design principles, most of these devices could be far more versatile at negligible, if any, increase in cost. They should also be more completely characterized in the literature so better informed decisions can be made as to the suitability of a specific device for a given application.

First, it would be useful to have a brief qualitative review of the requirements for making a crystal unit function in an oscillator and how to best utilize its unique properties.

A crystal controlled oscillator can generally be described as an amplifier with a feedback network containing a crystal resonator. To sustain oscillation, the loop gain must exceed unity at a frequency at which the total loop phase is $2n\pi$ ($n = 0$ or some whole integer). The level will be determined by the self-limiting characteristics of the loop or an external AGC circuit acting upon it.

Crystal controlled oscillators can generally be divided into two groups. Those that can close the phase loop with a resistor in place of the crystal are loosely called "series resonant". The crystal is inserted in the feedback loop to correct phase instability with a minimum of frequency change. (see figure 1) The amplifier may be noninverting. It may also be an inverter followed by a transformer, LC, or RC network or another active inverter.



Figure 1

The other type is an inverting amplifier using the crystal as a reactance (usually positive) in a phase inverting feedback network. This is commonly (and loosely) called a "parallel" or "anti-resonant" oscillator. (see figure 2)

If some frequency adjustment is desired on the first type of oscillator, a capacitor C_L can be inserted in series with the crystal. To close the phase loop, the crystal frequency will shift until its reactance cancels X_{C_L} . (see figure 3)

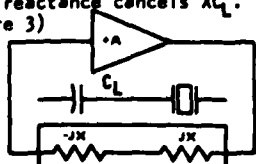


Figure 2

Figure 3

If the oscillator is ideal, the crystal in this case is not operating at series resonance. This is obvious to the crystal maker but not to many spec writers and circuit "designers".

In the case of the positive reactance oscillator, capacitors in the phase inverting network can be varied for frequency adjustment. (see figure 4)

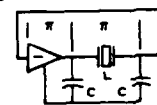


Figure 4

So far we have covered only the ideal case, 180° inverters or 0° amplifiers. With discrete circuits, the errors may be small enough to endure if not ignore in many applications. Also, the circuit designers generally had enough analog background to cope with such problems and the production feedback loop is fairly short so minor design corrections can be made as the project goes on.

With I.C.'s we have an entirely different situation. Most logic gates were not originally intended to be oscillators. They can be used for that function with varying degrees of success provided they are properly characterized. There is nothing inherently wrong with using a device for a purpose for which it was never intended, better known as improvising. There are two paths to success in such an endeavor. They are: 1: blind luck, 2: knowing what is inside the "black box" and what to do about it.

No further comment will be made on method 1 because everyone has their own system for dealing with it. Even method 2 suffers from the fact that a number of people receiving the same information, will not necessarily draw the same conclusions from it.

In the case of the non-inverting oscillator, usually made up of 2 inverters, especially in the case of I.C.'s where they are cheap and plentiful, the principal problem arises from two sources, phase error and phase instability.

- 1.) The inverter usually does not invert exactly 180° . For best stability the feedback loop should be capable of closing only thru' the motional arm of the equivalent circuit of the crystal unit. Unfortunately in most digital I.C.'s the phase error is in the wrong direction, making the oscillator appear inductive producing the situation shown in figure 7.

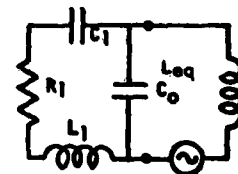


Figure 7

If the L_{eq} is very small so it resonates at a very high frequency with C_0 , it will cause only a slight lowering of crystal frequency and not much else. As it gets larger it resonates closer to the operating frequency, causing serious frequency shifts. Generally L_{eq} increases with frequency, adding to the problem. On some devices the frequency/inductance curve is complex as in figure 8.

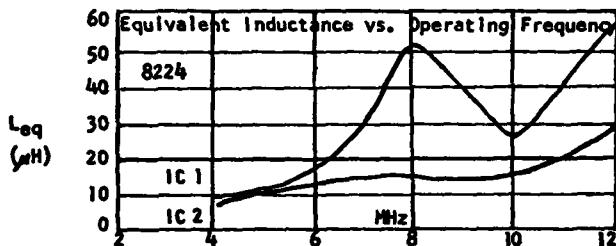


Figure 8

- 2.) Being made up of amplifier phase delay, L_{eq} is very unstable because it is a function of the instantaneous operating conditions on the amplifier elements. This is bad enough with circuits in which the capacitors and resistors are conventional components, but when they are made up of active devices the results can be many times worse. Active reactances also generally are less stable with temperature.

Both of these problems are compounded in the dual inverter circuit because the surplus of loop gain that is usually available makes it able to oscillate thru' most any feedback path. If L_{eq} is fairly large it may free-run thru' the pin to pin capacitance on the IC, the leads to the crystal unit or C_0 on a dead crystal.

In most of these circuits, when the frequency gets high enough so the gain is falling off enough to predict some improvement, the phase delay has also increased L_{eq} so the expected improvement does not occur. Some IC makers have sought the answer to this problem by having a computer design developed for a crystal unit that will not be so susceptible to frequency pulling by the oscillator. Unfortunately the resulting designs are not very producible.

When TTL first became popular, starting was a problem because the digital approach was to have an inverter either cut off or saturated. Many designers had no background in the conditions necessary for an oscillator to start. It took some time to convey the idea that it had to be biased in the linear region to start. Unfortunately on some of the earlier devices, variations from sample to sample and temperature made reliable starting a serious question.

For the oscillator to start, the loop must contain sufficient noise at the proper frequency and hopefully not too much at some other frequency where the gain-phase loop could close. Fast switching is not the secret of success in this case, even if it is the "in" thing in digital circuits because most of the unwanted responses of a crystal unit are higher in frequency than the desired ones.

With flexure mode units, usually below 100 KHz, the extension mode can start faster than the desired mode, in some cases preventing the desired mode from starting, making it shift after starting or in some cases, operating on both modes simultaneously.

Another frequent problem in the range of square DT cuts is the strong extension mode at 2.28 times the main mode. Except near the upper frequency limit of most gate type oscillators, they often start easier at the higher frequencies.

The fast wavefronts also represent energy at frequencies at which the reactance of C_0 may be low enough to cause trouble, even with fairly low L_{eq} . One of the most popular TTL gate oscillator circuits is shown in figure 11.

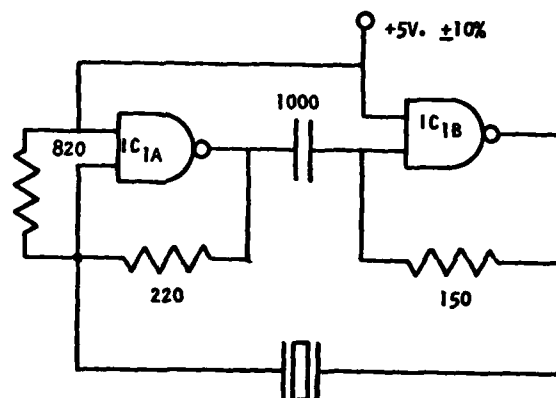


Figure 11

If the inverters were ideal it could be considered a series resonance oscillator and putting a load capacitor in series with the crystal would make the crystal operate as a positive reactance in which case it would not be a series resonance oscillator. In the usual gate IC the phase shift in an inverter is not 180° so the oscillator looks inductive at its terminals as shown in figure 12.

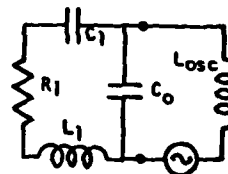


Figure 12

In this case a capacitor must be put in series with the crystal to bring it to series as shown in figure 13.

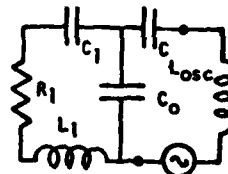


Figure 13

With an appropriate capacitor range, the crystal can be tuned either side of series resonance.

At first glance this looks like an ideal situation but it is not, for a number of reasons. The principal problem is the instability and scatter of the oscillator inductance. Also, the inductance usually increases with increasing frequency, so the capacitance required to cancel it would rapidly approach C_0 and may be even smaller.

A good example of this problem is shown in figure 14 on the 8224 which is promoted as the standard clock oscillator for the 8080 microprocessor.

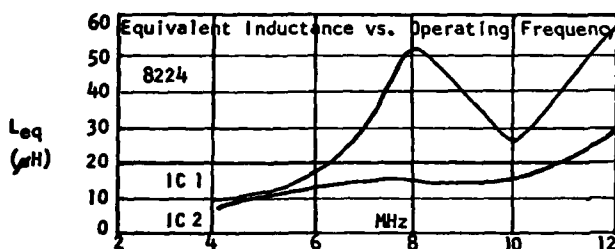


Figure 14

The next two slides show what happens to crystal performance with this type of device and what production spread from one manufacturer looks like. Some manufacturers hold better tolerances than others but the data sheets do not specify the factors critical to oscillator stability or operation.

8224 IC1

f_s (MHz)	Δf (Hz)	Δf (PPM)	L_{eq} (μH)	* C_L To Cancel L_{eq} (pf)
2	OPERATES ON 3RD. OVERTONE			
4	-156	-39	6.8	233
8	-2958	-370	13.8	29
10	-6776	-678	18.1	14
12	-12200	-1017	28.2	6
16	-59318	-3637	—	—
18	NOT CRYSTAL CONTROLLED			

8224 IC2

f_s (MHz)	Δf (Hz)	Δf (PPM)	L_{eq} (μH)	C_L To Cancel L_{eq} (pf)
2	OPERATES ON 3RD. OVERTONE			
4	-214	-53	9.4	170
8	-7455	-932	50.8	8
10	-8506	-851	25.2	1
12	-16503	-1376	54.8	3
16	-171947	-10541	—	—
18	NOT CRYSTAL CONTROLLED			

Figure 15

The data sheet on the 8224 does acknowledge a "small problem" with the following footnote. "When using crystals above 10 MHz, a small amount of frequency trimming may be necessary. The addition of a small capacitance (3 to 10pf) in series with the crystal will accomplish this function". In view of the data shown, this seems to be understated.

They also sanction overtone operation with the statement that overtone mode crystals generally have much lower "gain" than the fundamental type so an external LC network is necessary to provide additional gain for proper operation. Experience with this device indicates such operation is precarious at best. With 2 MHz crystals it operates only on the 3rd overtone at 6 MHz because the waveform has so much more energy at the higher frequency.

Some manufacturers of the 8224 have found that it is possible to improve these shortcomings somewhat with only minor changes, but it is still suitable only for loose tolerances.

Improvements requiring drastic changes in IC's are very expensive and take many months to put into production. In the meantime large quantities of the existing devices are in circulation, so the only solution is to provide adequate and accurate information on their limitations. This should induce the IC makers to do their design work and data sheets more thoroughly. Many of these problems could have been designed out initially at minimal cost compared to the detours to which their customers are unexpectedly subjected.

Another popular clock oscillator is the 12060L. Fortunately the manufacturer has published considerable information on its behavior as an oscillator but figure 18 presents more useful information and represents the average of several units. The production spread is not as bad as many other devices. This device gets into serious problems at a fairly low frequency. For higher frequencies the 12061L is suggested. Its behavior with frequency is shown in figure 19.

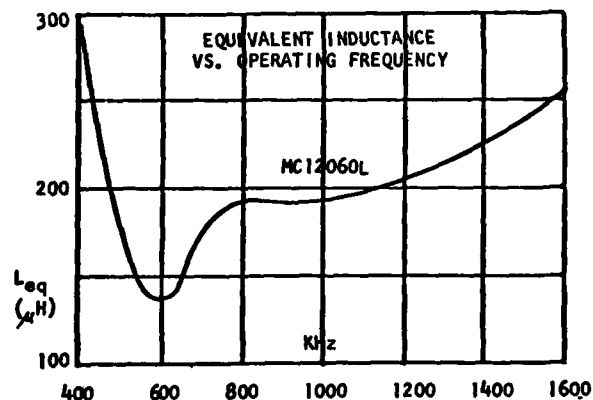


Figure 18

* C_L is capacitance required to cancel L_{eq} , bringing circuit to series resonance. Not to be confused with C_0 , which is to raise frequency above series resonance.

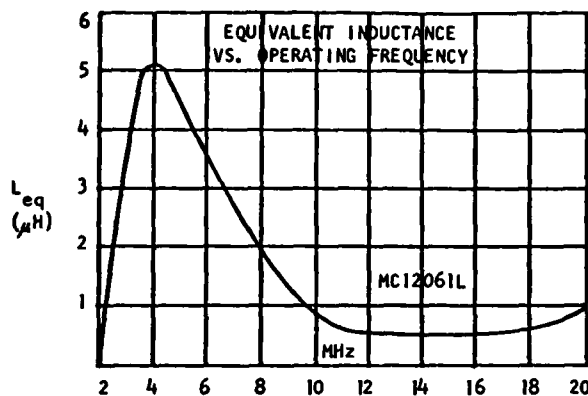


Figure 19

Note the peak at 4 MHz. It can be seen from this that measurements made only around 2 MHz or over 8 MHz can give an unrealistically optimistic picture if any part of the application is near 4 MHz.

The 14410 MOS series is an example of a good device with a deficient data sheet. For example, for 1 MHz it says: Crystal Mode - parallel. The crystal mode is fundamental, thickness shear. It operates in what is loosely called a parallel circuit but nowhere does it give the load capacitance, which is important to frequency calibration. It does give C_0 as "7pf typical" which is doubtful. Few 1 MHz crystal units in the HC-6/U enclosure run much over 3pf. Electrodes large enough to provide 7pf at 1 MHz would barely fit in the can.

No mention is made of the fact that an improvement can be made in stability on this unit by adding capacitors from each side of the crystal to common, and the resulting effects. While this may be of little consequence in the application for which it was intended, it is often put to other uses and such basic data should be provided.

Another device that needs better published data is the 74S124 family which is basically a VCO but with a data sheet that suggests its use as a crystal controlled clock oscillator as well. Results are shown in figure 22.

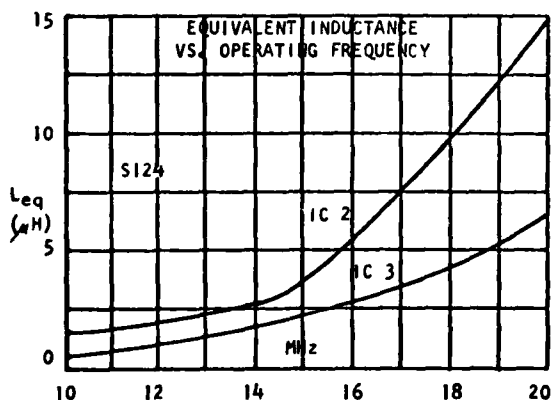


Figure 21

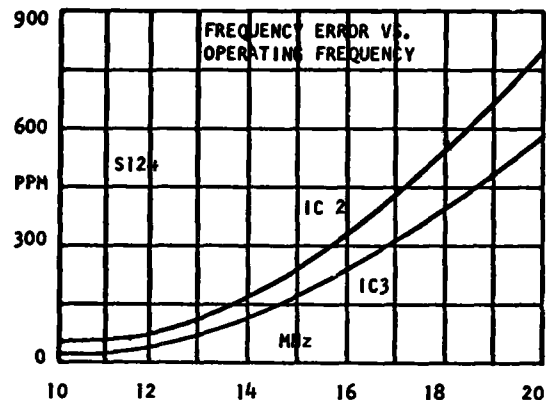


Figure 22

A further problem common to the 124 as well as the 4201 and S1856 is severe frequency shift with changes of crystal resistance. None of these effects are listed in the literature.

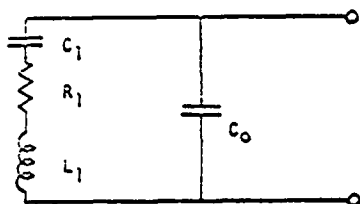
The data presented so far describes the primary problems with these relatively new devices. Some of the attempted solutions (with varying degrees of success or failure, depending on point of view) provide an interesting commentary on the clash of outlook between people of varying technical vintage.

Take for example, the designer who needs a calibration tolerance of ± 100 PPM for his system, so he orders crystal units to a ± 50 PPM tolerance and plugs them into a device that gives a production spread of ± 500 PPM. The obvious solution is to tighten the crystal tolerance - but the question is "How much"? There is a basic lack of understanding of how a crystal or any other tuned circuit controls frequency.

Some crystal manufacturers attempt to simplify information exchange with customers by publishing a list of crystal unit part numbers versus IC types for which they are recommended. For the devices with reasonable characteristics and systems with wide tolerances this has some virtue. It also has some drawbacks. It gives no clue as to the performance that can be expected, nor does it reveal the magnitude of the differences between types which might allow a different choice of IC if the first one turns out to be inadequate. It also implies that if you order that part number crystal unit for that IC, your troubles are over, when they have only begun.

The whole IC data situation needs drastic upgrading. One of the best is the RCA Application Note ICAN-6539. While it has some minor errors, if they were all that thorough, the savings of time and money would be well worthwhile. It was done by an engineer who did his homework, not by an advertising agent or marketing department clerk. In the words of an old Eimac satire vacuum tube data sheet, "Tests conducted in our sales department show that the performance is twice as good as you will find it to be".

Much confusion is generated by new and original names or abbreviations for well known terms. This is a quick way to identify questionable data. If the writer is not familiar with the terms he is using, the information has a high probability of being wrong, incomplete or misleading. A prime example is such terms as C_s or C_m for the motional capacitance of a crystal resonator. IEEE and IEC standards list it as C_1 and for good reason. Original abbreviations often turn out to have other meanings already established. For example C_m is modulus of compliance.



- f_s = FREQ. OF RESONANCE OF SERIES ARM
- f_m = FREQ. OF MIN. IMPEDANCE
- f_r = FREQ. OF ZERO REACTANCE AT RESONANCE
- f_a = FREQ. OF ZERO REACTANCE AT ANTI-RESONANCE
- f_n = FREQ. OF MAX. IMPEDANCE
- f_p = FREQ. OF PARALLEL RESONANCE

Quite a number of crystal manufacturers make the same error. Near the back of the Symposium proceedings is a list of the standards with which specification and technical data sheet writers and engineers should be familiar. They give terms and definitions as well as abbreviations.

Another area that needs attention is drive level. It may or may not be a problem in a given application. The number of crystal specs for IC's or IC data sheets that have obviously taken their drive level figures from MIL-C-3098 is an indication that the people involved have not determined the level to be expected in their circuits. In many cases they represent the state of the art in an era when crystal oscillators produced watts of output from a 300+ volt supply and the transistor had not yet been developed. They are not necessarily appropriate for the latest devices of today. A missing drive level spec is better than an erroneous one.

The "engineer of yesteryear" also discovered the fact that at some frequencies capacitors are inductive, inductors are capacitive and resistors can be either or both - and active devices all 3, plus negative resistance. Thus, stray resonances could turn up unexpectedly and usually did. These problems were resolved in the process of putting the components together into a specific device for a specific purpose. If one component resulted in a problem in a given situation, it would be replaced in minutes, and a design change was complete except for the paperwork.

With the IC this process has been moved to a stage where changing one unsuitable component in a device for one application is expensive, takes several months and may make it unsuitable for another application into which it has already been successfully designed.

In the case of devices developed and used for a specific purpose, it is difficult to rationalize some of the designs now in use, especially without the limitations carefully spelled out in the application data. Some of the worst examples are casually mentioned, as if they were no real problem at all.

One example of the difficulty this can create is as follows: Microprocessors are a convenient tool for handling the complex combinations required in frequency synthesizers. The microprocessor system has its own clock oscillator, so why not use it as the time base for the frequency synthesis? After the system is designed is an expensive time to discover that the clock oscillator stability of $10^{-5}/^{\circ}\text{C}$ is not satisfactory for the carrier on a single side-band system.

There are some very good devices on the market in which the design was done or assisted by people acknowledged in oscillator theory and requirements. These generally are also better characterized in the data.

In summary, a quartz crystal resonator is an analog and not a digital device. To use it with a digital device, a number of factors must be taken into account that can be treated more lightly or ignored when using only digital devices.

To arrive at a given frequency, the phase shift between the oscillator terminals must be considered an integral part of the frequency determining network. In order to do this, the device must be properly characterized. Ideally this should be done by the IC manufacturer in the configuration recommended and become part of the published data. Above all, it should be correct. The IC maker should also realize that there are limits to the range of attainable crystal parameters just as there are limits to the range of semiconductor parameters. The data should also include signal levels at the crystal terminals as well as a bibliography of good references.

Certainly the technical facts presented here are well known in the crystal industry. What is not so well appreciated is the magnitude of the problems described and how rapidly they are proliferating. The digital industry has given us many marvelous new developments but the rapid expansion of digital knowledge has in some cases obscured the principles upon which it was built. If analog and digital people are to accomplish things together, each of us must learn some of the other language or find a good translator.

References

1. Technical Report ECOM-2617 The Effect Of Load Capacitors On The Frequency Of Quartz Crystals. Arthur D. Ballato
2. Components Bulletin No. 6, Guide to the use of Quartz Crystals for Frequency Control, EIA
3. IEEE 177, The Piezoelectric Vibrator; Definitions and Methods of Measurement
4. IEC Publication 122.2, (International version of EIA Bulletin No. 6.)
5. IEC Publication 302, Standard Definitions and Methods of Measurement for Piezoelectric Vibrators Operating Over the Frequency Range up to 30 MHz.
6. IEC Publication 444, Basic Methods For The Measurement of Frequency and Equivalent Series Resistance of Quartz Crystal Units by Zero Phase Technique In A π -Network.
7. Analysis and Design of Crystal Oscillators, AD611-200, ECOM 2474. Eric Hafner. Available from: Defense Documentation Center, Attn: DDC-TCA, Cameron Station, Bldg. 5, Alexandria, VA 22314
8. Radio Engineering, Terman (McGraw Hill)
9. Electromechanical Transducers and Wave Filters, Warren P. Mason. Pub. by D. Van Nostrand Co.

THE MXO - MONOLITHIC CRYSTAL OSCILLATOR

T. LUXMORE AND D. E. NEWELL
CTS KNIGHTS, INC.
SANDWICH, ILLINOIS

Summary

A new type of hybrid circuit crystal controlled clock oscillator has been developed. The foundation for this approach has been provided by the efforts of W. Shockley, W. D. Beaver, W. L. Smith and others as presented in previous frequency control symposia.

The circuit essentially consists of a logic gate biased in its linear region (which serves as the gain element) and a monolithic coupled dual resonator for the frequency selective feedback path. The device is called a monolithic crystal oscillator (MXO). The MXO utilizes hybrid circuitry resulting in a new type oscillator which equals the performance of existing clock oscillators in general, surpasses their performance in some areas and remains very economical to construct. This paper discusses the design considerations and presents test data from several groups of MXOs.

REVIEW OF COUPLED RESONATOR

The crystal designed for this oscillator is an acoustically coupled dual resonator. It basically consists of two single resonators on a quartz wafer, with a coupling coefficient which is determined primarily by the plating configuration. Figure 1 shows the amplitude and phase versus frequency plot of a crystal near the standard IF frequency of 10.7 MHz with a bandwidth between the symmetric and antisymmetric modes of approximately 12 KHz. (For this application the bandwidth is not a critical parameter). As one can see the crystal displays 180° of phase shift at the symmetric point of minimum attenuation and 0° at the antisymmetric point. It is obvious that either an inverting or non-inverting amplifier could serve as the gain element in completing an oscillator circuit. One should also note that the source and load impedance to the resonator should be kept at a minimum to preserve the circuit response Q, since they are directly in series with each resonator of the coupled pair.

THE CIRCUIT

For our purpose a Schottky TTL NAND gate was selected as an inverting amplifier element. The remainder of this paper will be restricted to that device although many other inverting and noninverting gates available could be considered. The Schottky NAND gate was chosen for several reasons. It possesses a small propagation delay, the input and output impedances are relatively low and it enjoys a wide market resulting in a low purchase cost for the device.

The propagation delay of the device is an important parameter since it is a determining element in the oscillator's frequency stability. The slope of the resonator phase versus frequency curve is steepest at the 0° and 180° points. Any device delay shifts the operating frequency lower and into a portion of the phase curve with a lower slope, deteriorating the stability of the unit. In addition, if the gate displays a large propagation delay it will generally have large variations in that delay as a function of such variables as supply voltage, temperature, and time. This will produce shifts in phase at the initially lower slope point in the curve. Obviously this results in major fluctuations in frequency as a function of the variables mentioned. The Schottky NAND gate typically exhibits only 4 nanoseconds of delay.

Since low input and output impedances are important in preserving the response Q, the Schottky gate receives another plus. The output impedance of the device in this application is typically less than 50 ohms. The input impedance, though highly nonlinear and thus a function of drive amplitude, is roughly 300 ohms and slightly capacitive for the levels which normally occur in this application.

The method of linearizing the gate consists of a negative feedback bias utilizing two resistors (see Figure 2). The value of R2 is chosen to insure that the gate will remain in its transition region. The purpose of R1 is to establish a DC level at the gate input which will provide 50/50 symmetry at the output. The circuit parameter sensitivity to these resistors is low resulting in loose tolerances for both. To maintain a plus or minus 5 per cent symmetry band R1 must be maintained within 10 per cent and R2 within 20 per cent.

The gain and phase versus frequency characteristics have been empirically determined from evaluation of many 74S00 gates biased in the above manner. Figure 3 displays the typical response seen. The gain curve data was taken by driving the gate amplifier at low amplitude input levels which did not drive the device into a nonlinear output mode. The phase data was determined at a drive level which normally occurs during oscillation. Source and load impedances to the device were 50 ohms in both cases. From data taken over a 6 month period, it appears that the maximum deviation from this plot is typically 25 per cent for the gain response and 30 per cent for the phase curve.

TEST DATA

At this time many MXO's have been manufactured in both coldweld TO-8 and ceramic dual-in-line packages. The total circuit consists of the oscillator previously described and a gate connected to the oscillator output for isolation. Hybrid thick-film construction techniques were used in that the resistors and interconnections were screen printed onto a substrate and a 74S00 IC was epoxy bonded to the substrate and electrically connected via wire bonds. An open crystal was then mounted onto the substrate and final adjusted in the circuit with vacuum deposition techniques. The seal was accomplished with a coldweld process for the TO-8 and a ceramic cover was epoxied to the dual-in-line unit.

Several units were pulled at random from the manufacturing line and the data was collected which is presented here. Figure 4 exhibits the mean, standard deviation and range for a frequency versus supply voltage test. The upper set of data represents the maximum deviation from the nominal frequency at 5.0 volts. The lower set is the total change in frequency from 4.5 to 5.5 volts. Figure 5 displays the statistics of rise time and fall time tests (between the .4 volt and 2.4 volt levels) as a function of temperature variation. Figure 6 shows duty cycle data as a function of both supply voltage and temperature. Figure 7 presents the start-up characteristics of the oscillators. The first portion of data was collected from a measurement of time delay between the application of a step in voltage (from 0. volts) and the presence of a full waveform at the output terminal. The second portion of the data represents the supply voltage at which the unit first begins oscillating when a slow ramp is applied to the supply terminal. Figure 8 exhibits frequency versus temperature data. Obviously this is primarily a function of the particular cut of the crystal. Actual contribution of propagation delay shift was measured by isolating the crystal in a stable temperature environment and varying the temperature of the remainder of the circuit. Typical frequency shift from 0°C to 70°C at 10 MHz was found to be less than 3 parts per million (ppm). Aging data is shown in Figure 9 for units assembled in a coldweld TO-8 package. The rates are primarily a function of sealing processes but delay shift with time was also a concern. Several of these units were opened at the conclusion of this test and propagation delay was found to have shifted a maximum of 6 per cent. Power drain information was also gathered over a period of time from over 1000 units. Typical supply current was 16 milliamperes (ma) and the absolute maximum drain from the supply at 5.5 volts was 30 ma.

CONCLUSION

In conclusion, the authors believe that the MXO, with a total of only 4 components, the crystal cost not being significantly higher than standard single resonators of similar quality, is a very economical design to manufacture. Also the design maintains high standards of performance relative to most TTL system applications.

BIBLIOGRAPHY

- Beaver, W.D., "Theory and Design of the Monolithic Crystal Filter", Proceedings of the 21st Annual Frequency Control Symposium, Fort Monmouth, New Jersey, US Army Electronics Command, 1967, pp179-199
- Johnson, R.A., "Mechanical Bandpass Filters": Temes and Mitra, (editors); Modern Filter Theory and Design, New York, Wiley, 1973, Chapter 5, pp. 157-210
- Shockley, W., Curran, D.R., and Koneval, D.J., "Energy Trappings and Related Studies of Multiple Electrode Filter Crystals", Proceedings of the 17th Annual Frequency Control Symposium, Fort Monmouth, New Jersey, US Army Electronics Research and Development Laboratory, 1963, pp.88-126
- Smith, W.L., "The Application of Piezoelectric Coupled-Resonator Devices to Communications Systems", Proceedings of the 22nd Annual Frequency Control Symposium, Fort Monmouth, New Jersey, US Army Electronic Command, 1968 pp. 206-225

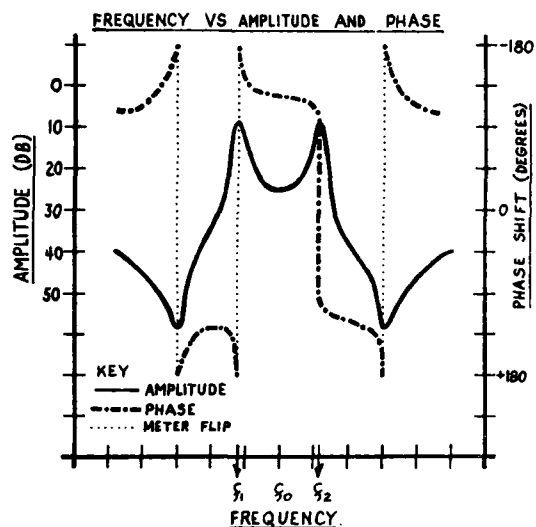


FIGURE 1 FREQUENCY VERSUS AMPLITUDE AND PHASE FOR A DUAL RESONATOR

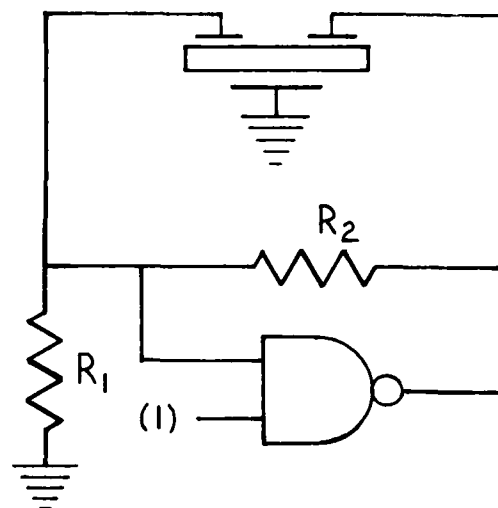


FIGURE 2 MXO SCHEMATIC

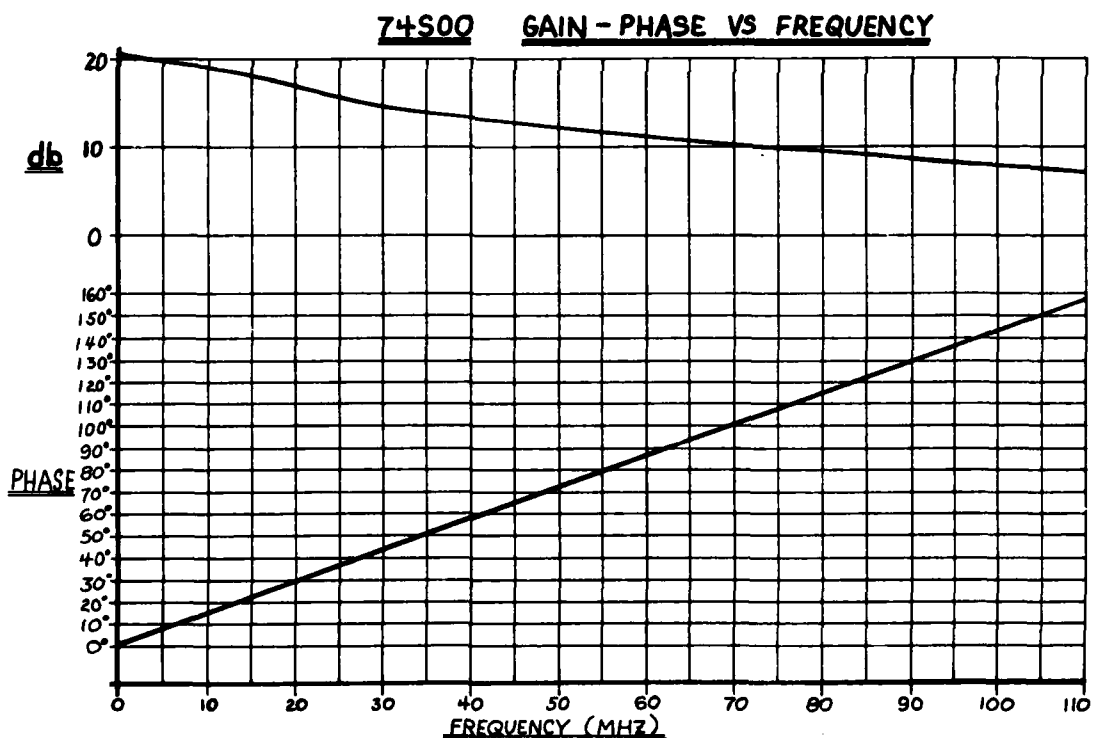


FIGURE 3 GAIN-PHASE VERSUS FREQUENCY OF A LINEARIZED 74S00

VOLTAGE COEFFICIENT FOR 143 UNITS

CONDITIONS: ROOM TEMPERATURE, 54S00,
10 STANDARD TTL LOADS

WORST CASE

5 VOLTS $\pm 10\%$ $\bar{x} = 0.750$ PPM
 $\sigma = 0.44$ PPM
RANGE = 0.203 TO 2.035 PPM

OVER 4.5 TO
5.5 VOLT RANGE $\bar{x} = 1.055$ PPM
 $\sigma = 0.61$ PPM
RANGE = 0.203 TO 2.035 PPM

FIGURE 4 VOLTAGE COEFFICIENTS

DUTY CYCLE (SYMMETRY) FOR 143 UNITS

CONDITIONS: VOLTAGE 5 VOLTS $\pm 10\%$, 54S00,
10 STANDARD TTL LOAD

-55°C: $\bar{x} = 50.48\%$
 $\sigma = 2.44\%$
RANGE = 48 TO 52.5%

0°C: $\bar{x} = 51.24\%$
 $\sigma = 1.14\%$
RANGE = 49 TO 54%

80°C: $\bar{x} = 51.49\%$
 $\sigma = 0.99\%$
RANGE = 49 TO 54%

FIGURE 6 DUTY CYCLE

TEMPERATURE TESTS FOR 72 UNITS

CONDITIONS: 5 VOLT SUPPLY, 74 S00

0°C TO 80°C: $\bar{x} = 13.22$ PPM

$\sigma = 7.55$ PPM

RANGE = 2.6 TO 29.5 PPM

-55°C TO 80°C: $\bar{x} = 22.78$ PPM

$\sigma = 14.26$ PPM

RANGE = 2.0 TO 48.4 PPM

FIGURE 8 TEMPERATURE COEFFICIENTS

RISE AND FALL TIME FOR 143 UNITS

CONDITIONS: -55°C, 0°C AND 80°C, 54S00,
10 STANDARD TTL LOAD,
FREQUENCY 8-24 MHz

	-55°C	0°C	80°C
RISE TIME: \bar{x}	3.64 NS	2.72 NS	2.90 NS
σ	0.15 NS	0.49 NS	0.49 NS
RANGE	3.4 TO 3.9 NS	2 TO 4 NS	2 TO 4 NS
FALL TIME: \bar{x}	2.92 NS	2.92 NS	2.92 NS
σ	0.20 NS	0.33 NS	0.64 NS
RANGE	2.7 TO 3.2 NS	1 TO 3 NS	2 TO 4 NS

FIG. 5

FIGURE 5 RISE AND FALL TIMES

STARTING TIME AND VOLTAGE FOR 16 UNITS

0°C TO 80°C, 54S00, 10 STANDARD TTL LOAD

VOLTAGE	4.5	5.0	5.5
MAX. STARTING TIME	1.0 MS	0.8 MS	0.2 MS
STARTING VOLTAGE	< 3.5 VOLTS		

FIGURE 7 START-UP CHARACTERISTICS

AGING DATA (40 UNITS)

CONDITIONS:

PACKAGE TO-8

TIME 10,000 HRS

TEMP. 85°C

$\bar{x} = 4.5$ PPM

$\sigma = 5.8$ PPM

RANGE = 17.4 PPM

FIGURE 9 AGING DATA



FIGURE 10 TO-8 AND DIP MXO'S

LINEAR CRYSTAL CONTROLLED FM SOURCE FOR
MOBILE RADIO APPLICATION

by

R. Arakelian
Iran Electronics Industries

and

M. M. Driscoll
Westinghouse Electric Corporation

Summary

In connection with a VHF mobile radio design effort at Iran Electronics Industries, a narrow deviation voltage controlled crystal oscillator (VCXO) circuit has been developed that exhibits near linear tuning characteristic and, unlike previously reported circuitry, is extremely simple and straightforward in design and highly economical with regard to material and fabrication costs.

The design principle makes use of a linear reactance modulator incorporating abrupt junction varactor diode developed by M. Driscoll and previously reported by Driscoll and Healey in connection with the design of a wide deviation VCXO for radar application. Because the total oscillator tuning requirements are only on the order of 1 part in 10^5 however, much of the circuit complexity has been eliminated, and the problem of interdependence between linearity and center frequency adjustment has been solved. Network equivalence theorems have been utilized to reduce the total number of required circuit inductances to only one, and the composite resonator has been re-configured to operate on the series resonance of its impedance characteristic allowing simple means for prevention of oscillation at spurious network resonances through the addition of a single resistor.

Laboratory measurement of transmitter FM distortion using a number of FM oscillator modules reveals repeatable achievement of less than 0.5% distortion with linearity adjustment and less than 1% without any adjustment. Also, the achievement of repeatable linear tuning characteristic has simplified the oscillator modules so that required temperature stability is obtained by addition of MTC capacitors to the oscillator circuit and a single simple 3 point thermistor compensation circuit that is easily designed and provides simultaneous temperature dependent compensating voltage for all radio oscillator modules. The oscillator design also allows for direct parallel connection of multiple modules (for multichannel operation) having commonly connected tuning input and signal output ports.

Introduction

The upper VHF band allocated for mobile radio communications covers the frequency range 132 to 174 MHz with ± 5 KHz deviation standard for narrow band FM systems.¹

This deviation corresponds then to ± 29 to ± 38 ppm transmitter carrier shift. Minimum standards established for VHF mobile transmitters specify 10% maximum modulator distortion and 0.0005% maximum carrier drift over an operating temperature range of -30 to 60°C .²

Carrier long term stability requirements are such that the transmit (and receiver local oscillator) frequencies must be crystal controlled and that either temperature compensation or control of the crystal unit(s) is required. Largely due to size and power consumption considerations, temperature control is an unacceptable alternative, and temperature compensation is employed even in less expensive commercial transceivers.

The most common means employed for temperature compensation of the quartz crystal unit involves modification of the crystal oscillator circuit for voltage tuning, coupled with the use of thermistor networks designed to provide a temperature dependent voltage characteristic that "tunes" the oscillator so as to cancel its frequency drift with temperature.

It would appear that since voltage tuning of the oscillator is required for temperature compensation, this capability could also be utilized for required audio modulation of the transmit carrier frequency as well. This is usually not the case with transceivers on the market today however, because previously reported means for linearization of the oscillator tuning characteristic have proven too complex to be cost effective. Instead, phase modulation of a fixed frequency carrier is employed. Due to the high modulation index required at low audio rates (even at sub-multiples of the final transmit frequency), it becomes necessary to use several cascaded phase modulators, each providing typically less than one radian peak phase excursion. Also, de-emphasis circuitry (in addition to the usual pre-emphasis plus limiting circuits) must be included for simulation of true FM. Clearly, if the problem of inexpensive means for linear crystal oscillator tuning could be solved in a way that would not degrade oscillator frequency stability or increase oscillator size and complexity, direct FM as well as temperature compensation could be achieved with significant reduction in overall transceiver cost and complexity.

Background

In 1967, one of the authors devised a scheme for the design of a crystal-controlled voltage-tunable, composite resonator that exhibited near linear relationship between applied tuning voltage and composite resonator resonant frequency. The resonator design was based on the development of a linear reactance modulator incorporating abrupt junction varactor diodes. The main objective of that effort was directed by Messrs. Driscoll and Healey towards the successful development of a wide deviation voltage controlled crystal oscillator (VCXO) capable of 0.1 to 0.5 percent tuning for radar application.³ The resulting circuit, while highly successful, was more complex and costly than the type of oscillator circuitry employed in modern mobile transceivers. Much of the circuit complexity resulted however, from the severe performance requirements, including the requirement for relatively large tuning range.

In connection with a mobile transceiver development program conducted at Iran Electronics Industries, the potential for use of the linear reactance modulator concept in the design of a simpler, narrow deviation VCXO was investigated.

Linear, Narrow Deviation VCXO Composite Resonator

Figure 1 shows the reactance versus voltage characteristic exhibited by an abrupt junction varactor diode whose capacitance is given by:

$$C_2 = KV^{-\gamma} \quad (V = V_a + \phi) \quad (1)$$

where V = the applied tuning voltage, V_a plus the contact potential, $\phi \approx 0.6$ volts. For the abrupt junction diode, $\gamma \approx 0.5$. These diodes are usually specified in terms of the 4-volt (applied voltage) capacitance, where:

$$C_2 = C_4 \left(\frac{4 + \phi}{V} \right)^\gamma \quad (2)$$

Figure 2 shows the reactance versus voltage characteristic exhibited by a parallel diode/inductance combination. As can be seen from the figure, there is an inflection point in the characteristics (shown where the applied voltage = V_0), and the characteristic is near-linear in the vicinity of the inflection point. The expression for the reactance of the circuit of figure 2 is:

$$X = \frac{\omega L_2}{\omega^2 L_2 KV^{-\gamma} - 1} \quad (3)$$

Differentiating with respect to voltage:

$$dX/dV = K_2 = \frac{-\omega L_2 (\gamma \omega^2 L_2 KV^{-\gamma+1})}{(\omega^2 L_2 KV^{-\gamma} - 1)^2}$$

$$= \frac{-\gamma \omega^3 L_2^2 C_2}{(\omega^2 L_2 C_2 - 1)^2 V} \quad (4)$$

Letting $d^2X/dV^2 = N(V)/D(V)$, and setting $N(V) = 0$ (at the inflection point):

$$\omega^2 L_2 KV^{-\gamma} = \omega^2 L_2 C_2 = \frac{\gamma+1}{1-\gamma} (V=V_0 + \phi) \quad (5)$$

or, in terms of the 4-volt capacitance:

$$L_2 = \left(\frac{1}{\omega^2 C_4} \right) \left(\frac{1+\gamma}{1-\gamma} \right) \left(\frac{V_0 + \phi}{4 + \phi} \right)^\gamma \quad (6)$$

Setting $\gamma = 0.5$ and rearranging terms:

$$L_2 = \frac{3}{\omega^2 C_4} \sqrt{\frac{V_0 + \phi}{4 + \phi}} \quad (7)$$

Substituting the expression for L_2 in equation (5) into equation (4), we find the expression for the slope, K_2 , of the reactance characteristic at the inflection point:

$$\frac{dX}{dV} = K_2 = \left[-\frac{1}{\omega C_2} \right] \left[\frac{(1+\gamma)^2}{4\gamma} \right] \left[\frac{1}{V_0 + \phi} \right] \quad (8)$$

Again, equation (8) may be written in terms of diode 4-volt capacitance:

$$K_2 = - \left[\frac{1}{\omega C_4} \right] \left[\frac{(1+\gamma)^2}{4\gamma} \right] \frac{(V_0 + \phi)^{\gamma-1}}{(4 + \phi)^\gamma} \quad (9)$$

For $\gamma = 0.5$:

$$K_2 = - \frac{9}{8} \left[\frac{1}{\omega C_4} \right] \left[\sqrt{\frac{1}{(V_0 + \phi)(4 + \phi)}} \right] \quad (10)$$

Figure 3 shows the reactance versus frequency characteristic for the crystal unit with static capacitance, C_0 , antiresonated by means of a parallel inductor, L_0 . In the vicinity of crystal unit series resonance, the reactance characteristic is near linear and is given (approximately) by:

$$K_1 = 2L_1 \text{ ohm-sec/radian} \quad (11)$$

From figures 2 and 3 it can be seen that if the networks of figures 2 and 3 were connected in series, as shown in figure 4, the resulting network would exhibit near-linear relationship between series resonant frequency and tuning voltage (shown for a tuning voltage range of V_1 to V_2 and a frequency range ω_1 to ω_2 in figures 2 and 3, respectively), with overall resonance occurring slightly above crystal unit series resonance. The resulting tuning sensitivity given as:

$$K_0 = (\omega_2 - \omega_1) / (V_2 - V_1) \text{ radians/volt-sec} \quad (12)$$

We can also write:

$$K_2 = K_1 K_0 \text{ ohms/volt} \quad (13)$$

Substituting from equations (10 and (11), we can rewrite (13) as:

$$-\frac{9}{8} \left[\frac{1}{\omega_C^2} \right] \left[\frac{1}{(\sqrt{V_0 + \phi})(4 + \phi)} \right] = 2K_0 L_1 \quad (14)$$

or,

$$L_1 = \left(-\frac{9}{16K_0} \right) \left(\frac{1}{\omega_C^2} \right) \left(\frac{1}{(\sqrt{V_0 + \phi})(4 + \phi)} \right) \quad (15)$$

Equations (15) and (17) comprise the basic design equations for the narrow deviation VCXO composite resonator.

In an effort to minimize the number of circuit inductances (since they are usually the least stable, lossiest resonator components, and do not lend themselves to miniaturization), we can, as shown in figure 4, combine L_0 and L_2 without effectively altering the resonant frequency versus voltage characteristic of the dipole, as long as the series resonant frequency does not approach the parallel resonant frequency of the crystal unit. Figure 5 shows the reactance versus frequency characteristic for the circuit of figure 4(b). As can be seen from the figure, in addition to the desired resonance at ω_0 , there is an undesired or spurious resonance at ω'_0 below ω_0 . We will deal with the problem of suppression of this spurious resonance later in the paper.

The Sustaining Stage

Figure 6 shows, in generalized form, the connection of the composite resonator to the sustaining stage to form the oscillator. The sustaining stage acts as a negative resistance generator whose input impedance is given ideally by:

$$Z_{IN} = -j \left(\frac{1}{\omega C_5} + \frac{1}{\omega C_6} \right) \frac{1}{\omega^2 C_5 C_6 (1/g_m + R_E)} \quad (16)$$

The second term of equation (16) constitutes the negative resistance term. R_E constitutes a form of local negative feedback which may be included in the circuit in order to stabilize the negative resistance term (and thus the excess gain of the oscillator) against variations in g_m .⁴ Also shown in figure 6 is a variable reactance (C_3, C_4) which has been included in the circuit to provide a means for center frequency adjustment of the oscillator. The net reactance due to the center frequency and sustaining stage circuits "seen" by the composite resonator is capacitive and has negligible effect on the linearity or tuning sensitivity exhibited by the composite resonator, so long as this reactance is small compared to that of the crystal unit static capacitance. This is illustrated in figure 7. As shown in figure 7, the effect of a capacitive load is to change the apparent value of crystal unit motional inductance (proportional to K_1 and thus tuning sensitivity), as well as apparent static capacitance. Of

course, once C_3 through C_6 have been selected, their effect in modifying the apparent values of crystal unit parameters can be predicted and accounted for in the design of the composite resonator. It should be appreciated here that an additional inductor could be placed in series with the crystal unit to tune out the reactive portion of the sustaining stage input impedance, or, for that matter, of both the sustaining stage and modulator capacitive reactances (permitting operation about the series resonant frequency of the crystal unit). For reasons previously listed however, this was not done for our application.

As can be seen from figure 6, the crystal unit is now effectively "loaded" by a net capacitive reactance due to the sum of the reactances of the modulator, center frequency adjustment, and sustaining stage circuits. The equivalent load capacitance due to these is easily computed, and the crystal unit frequency (nominal oscillator frequency) is specified with the appropriate value of nominal load capacitance.

Design Example

Figure 8 shows the actual schematic diagram for an FM source designed to operate at 12.5 MHz with a nominal tuning sensitivity of 250 Hz/volt and a maximum tuning range of ± 500 Hz. As a starting point in the design procedure, the value for crystal unit motional capacitance of 0.02 pF was chosen. This type of crystal unit will have C_0 250 pF, C_1 5 pF, and a reasonable value for crystal unit series resistance lies in the range 15 to 25 ohms. L_1 for the crystal unit = 8.1057 mH, so that the reactance slope in the vicinity of series resonance = $2L_1 = 0.0162$ ohm-sec/radian. Choosing $C_4 = 82$ pF and $C_3 = 15$ -60 pF will therefore provide a reactance variation of 40 ohms for center frequency adjustment; more than enough to accommodate a ± 0.001 percent adjustment range. $C_5 = C_6 = 220$ pF and $R_E = R_5 = 22$ ohms (with $g_m \approx 0.2$ mhos at 5 mA transistor current) will provide a negative resistance of approximately 93 ohms. The oscillator excess gain will therefore be on the order of 10dB, assuming a composite resonator net series resistance of approximately 30 ohms (primarily due to losses in the crystal unit, plus additional loss introduced by R_6). R_6 is included in the circuit as an extremely simple means of preventing oscillation at the spurious resonance of the composite resonator (ω'_0 in figure 5). This can be explained as follows: Over the desired tuning range of the oscillator, the maximum reactance exhibited by the crystal unit is on the order of 500 ohms. With $R_6 = 33$ Kohms, the maximum equivalent series resistance due to R_6 is only 7 ohms. At the spurious resonance however, the reactance of the crystal unit is substantially that of its static capacitance (approximately 3700 ohms), and the equivalent series resistance introduced by R_6 is on the order of 400 ohms. Since this is much larger than the negative resistance generated by the sustaining

stage, oscillation at the spurious resonance is prevented.

The net capacitive reactance due to C_3 through C_6 is on the order of 225 ohms, corresponding to an equivalent capacitance of 56 pF. From figure 7 therefore, the apparent values for crystal unit motional inductance and static capacitance (as "seen" by the modulator) are 8.83 mH and 4.6 pF, respectively. Choosing a nominal varactor bias of $V_0 = 5.6$ volts, and solving equation (15) for C_4 :

$$C_4 = \frac{9}{16 (2\pi)^2 (-250\text{Hz/V})} \cdot \frac{1}{2\pi (12.5 \times 10^6) (8.83 \times 10^{-3})} \cdot \frac{1}{\sqrt{(6.2)(4.6)}} = 97\text{pF}$$

This is close enough to the standard value of 100 pF, so that a 100 pF diode is selected. Now, from equation (7),

$$L_2 = \frac{3}{4\pi^2 (12.5 \times 10^6)^2 (100 \times 10^{-12})} \sqrt{\frac{5.6 + 6}{4 + 6}} = 5.65 \mu\text{H}$$

The crystal unit apparent static capacitance is 4.6 pF. Therefore $L_0 = 35 \mu\text{H}$. L_2' is the parallel combination of $L_2 = 5.65 \mu\text{H}$ and $L_0 = 35 \mu\text{H}$ or $4.86 \mu\text{H}$. A molded choke having an inductance of $4.7 \mu\text{H}$ was used for L_2 .

Figure 11 illustrates the method selected for extraction of signal power from the oscillator module and also indicates the means of combining multiple modules (one per channel) having common signal output and tuning input ports. Signal current is extracted at the sustaining stage transistor collector terminal via a common base buffer amplifier exhibiting very low input impedance. In this manner, the voltage developed at the sustaining stage collector terminal is small, and feedback effects (via the transistor collector to base capacitance) due to load variations are minimized. The selection of relatively large excess gain in the sustaining stage results in an oscillator output current consisting of a small conduction angle sinusoid, so that extraction of the third harmonic of the fundamental oscillator operating frequency at the buffer output is possible. Subsequent multiplication by four in the transmitter and three in the receiver results in generation of final transmit and receiver local oscillator signal at VHF.

Temperature Compensation

The most suitable type of crystal unit for use in this application is the so-called AT cut whose temperature characteristic is shown in figure 12. As can be seen from the figure, for a ± 2 minute tolerance in the cutting angle, the frequency change with temperature will be on the order of ± 0.001 percent or double the maximum allowable drift in transceiver carrier

frequency. Because the temperature compensation requirements are so modest, a relatively crude and simple means can be employed for temperature compensation of the oscillator modules(s).

Table 1 indicates the various values for temperature coefficient required in C_5 and C_6 of figure 8 in order to obtain nominal zero temperature coefficient at room temperature in the oscillator frequency temperature characteristic. Figure 13 shows the resulting temperature characteristic, with the shaded area representing the region of uncertainty due to non-exact coefficient in C_5 , C_6 . Also, shown in the figure is the required three-section thermistor correction characteristic required for oscillator compensation at the low and high ends of the temperature range. This type of characteristic is extremely easy to achieve with the use of the circuit figure 14. In the circuit of figure 14, at the low temperature portion of the operating temperature range, Q_3 is biased off due to RT_2 , and the voltage-temperature response is determined by RT_1 . Over the mid temperature range, both Q_1 and Q_3 are biased off, and the output voltage is fixed at a level determined by R_3 through R_6 . Similarly, at the high end of the operating temperature range, Q_1 is biased off, and the circuit response is governed by RT_2 variation with temperature. Although the compensation circuit offers very coarse three section straight line compensation, it is extremely easy to design using two identical thermistors having non critical R-T characteristic and can be used to provide simultaneous compensating voltage for all radio oscillator modules.

Figure 15 shows the measured response for the circuit of figure 14, and the frequency-temperature characteristic measured for the uncompensated, partially compensated (C_5, C_6 compensation only), and fully compensated oscillator module is shown in figure 16. As can be seen from figure 16 compensation to ± 3 ppm was readily achieved using the technique previously described.

Conclusions

A procedure for the design of linear narrow deviation VCXO ideally suited for mobile transceiver application has been presented. The procedure makes use of a set of straightforward design equations for the design of the frequency determining portion of the oscillator circuit, and the interrelationship between circuit element values as indicated in the design equations allows the circuit designer to make intelligent tradeoffs in selection of component values in order to achieve required device performance. Although in the application described the number of circuit inductances was reduced to one, added flexibility can be achieved with the addition of a second inductor for achieving wider tuning ranges with operation nominally at crystal unit series resonance. Also, an ex-

tremely simple means for suppression of oscillation at composite resonator spurious resonances has been described.

The sustaining stage has been configured so as to minimize load variation effects on oscillator stability, and further allows for direct parallel connection of multiple modules having common signal output and tuning input ports.

Although the application described is for generation of direct FM in mobile transceivers, The device can find important uses in other systems requiring linear narrow deviation VCXO.

Bibliography

1. D. Hamsher, Communication Systems Engineering Handbook, McGraw-Hill, 1967, chapter 17.
2. EIA Specification RS152B, "Minimum Specifications for Land-Mobile FM or PM Transmitters". February, 1970.
3. Driscoll and Healey, "Voltage Controlled Crystal Oscillators", IEEE Trans. on Electron Devices, Aug., 1971 Vol. ED-18, No. 8, pp528-535
4. M. M. Driscoll, "HF and VHF Source Stabilization Incorporating Q-Multiplied Quartz Crystal Unit", IEEE Trans. on I and M., June, 1974, Vol IM-23, No 2, pp 131-140.

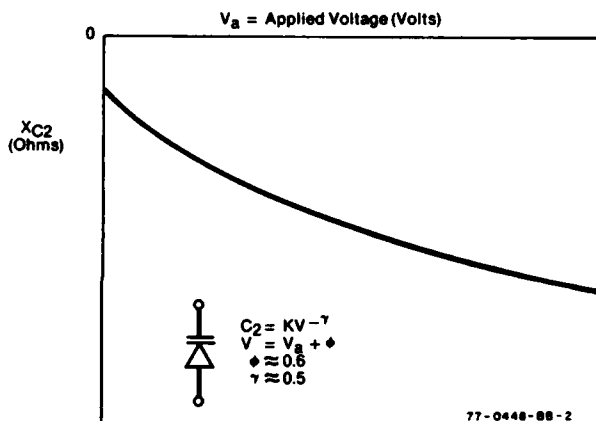


Figure 1. Reactance vs Voltage for Abrupt Junction Diode

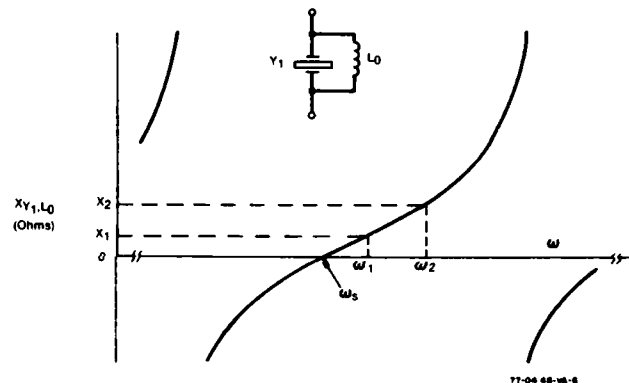


Figure 3. Reactance of Crystal Unit With Anti-Resonated Crystal Capacitance

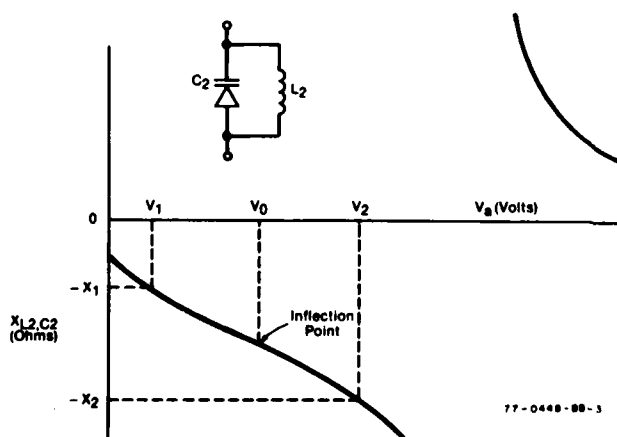


Figure 2. Reactance vs Voltage for Parallel Diode/Inductor

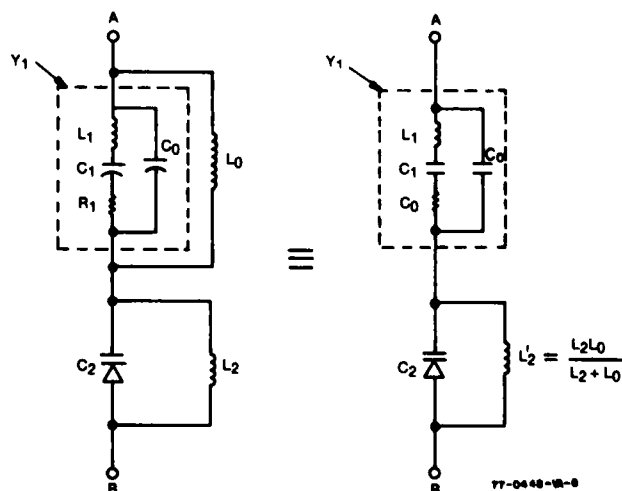


Figure 4. Combining Inductors In the Composite Resonator

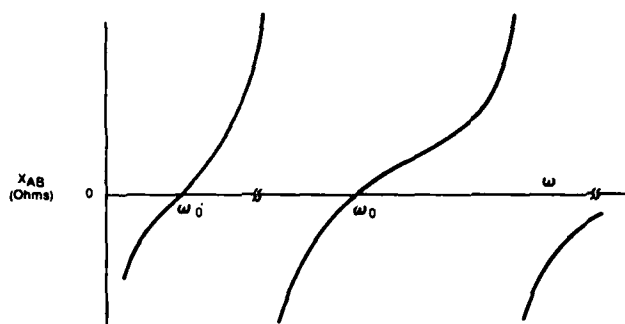


Figure 5. Reactance vs Frequency Characteristics for Composite Resonator

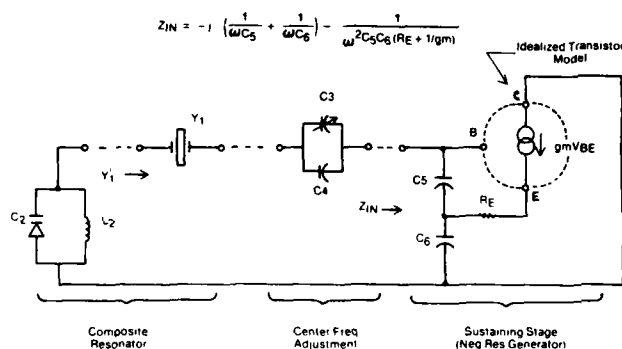
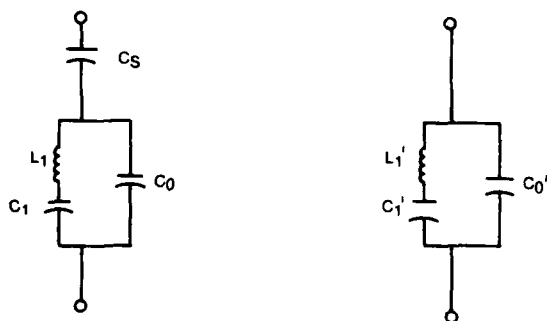


Figure 6. Composite Resonator Connection to Sustaining Stage



$$L_1' = L_1 \left(1 + \frac{C_0^2}{C_S} \right)$$

$$C_0' = C_0 C_S / (C_0 + C_S)$$

Figure 7. Effect of Series Capacitance On Apparent Crystal Parameters

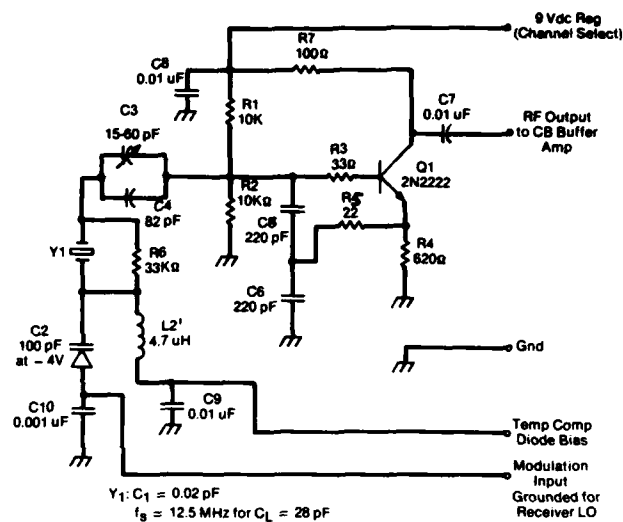


Figure 8. Transmitter (and Receiver) Oscillator Module Schematic Diagram

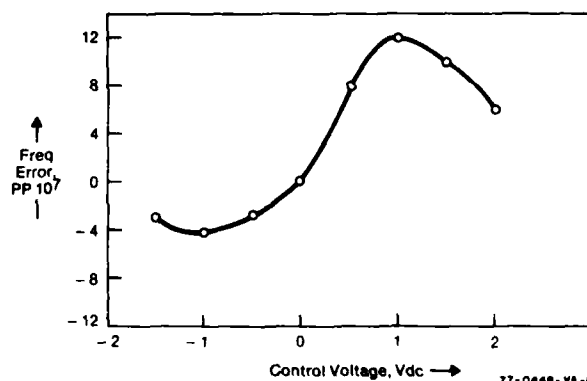


Figure 9. Error for Experimental Oscillator Modules

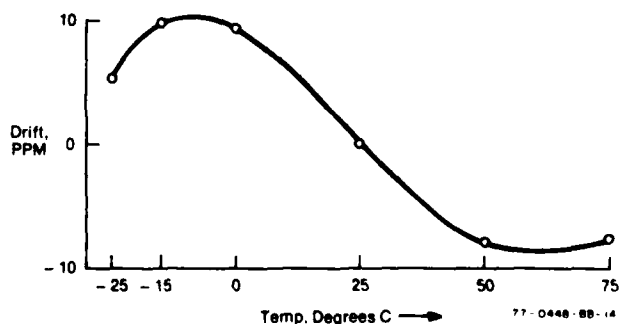


Figure 10. Frequency Temperature Drift for Uncompensated Oscillator Module

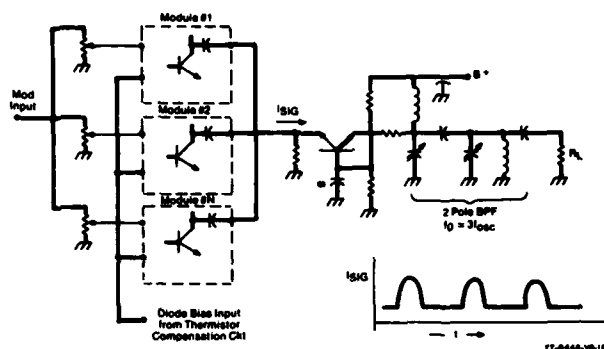


Figure 11. Oscillator Module Interconnection

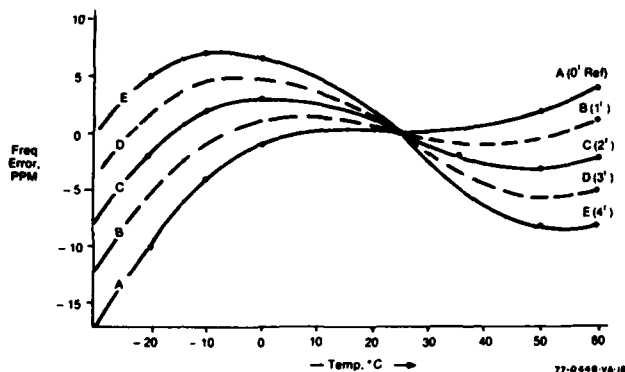


Figure 12. Frequency-Temperature Characteristics At-Cut Crystal Unit

TABLE 1
REQUIRED TEMPERATURE COEFFICIENT
IN C5, C6 FOR ROOM TEMPERATURE COMPENSATION
OF OSCILLATOR FREQUENCY

Curve No.	Crystal Unit T.C. at 25°C (PPM/°C)	C5, C6, T.C. (PPM/°C)	Standard T.C. Available (PPM/°C)
A	0	-750	-1000
B	-0.085	-1900	-2200
C	-0.17	-3000	-3300
D	-0.275	-4200	-4300
E	-0.34	-5300	-5600

L2
φ (in C2)
C3, C4
T.C. = 200 PPM/°C
T.C. = -2 mV/°C
T.C. = -150 PPM/°C

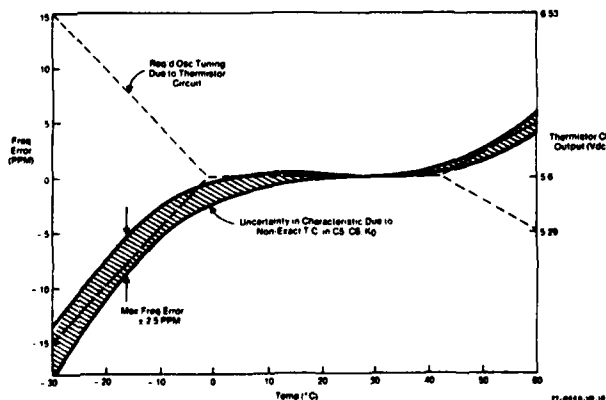


Figure 13. Graphical Representation Of Required Thermistor Circuit Response

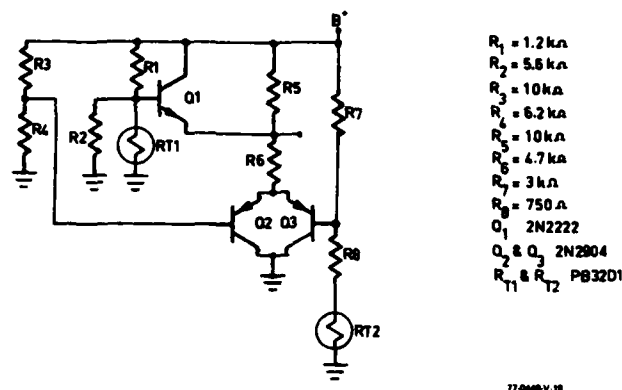


Figure 14. Temperature Compensation for IEI Radio Oscillator

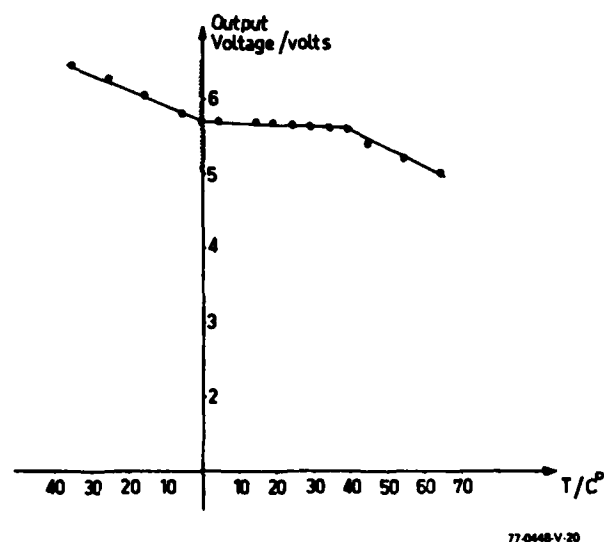


Figure 15. Thermistor Bridge Voltage vs Temperature

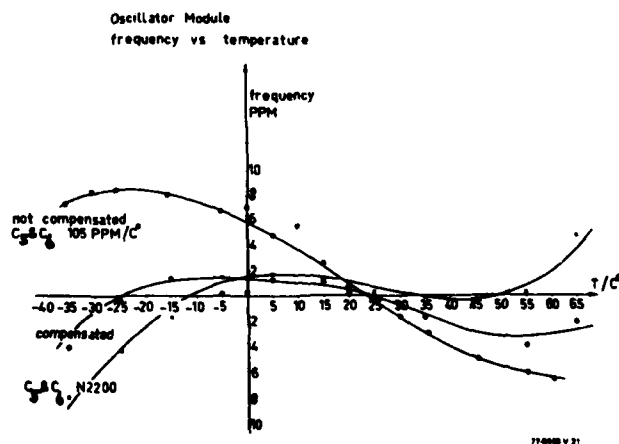


Figure 16. Oscillator Module Frequency vs Temperature

DESIGN CONSIDERATIONS FOR A
DIGITALLY TEMPERATURE COMPENSATED CRYSTAL OSCILLATOR

Paul J. Scott
Motorola Communications Group
Schaumburg, Illinois

Summary

Crystal oscillators to be used over extended environmental ranges must be compensated for temperature effects. Since these effects repeat in temperature, all that is needed is a circuit to correct the oscillator frequency and some form of memory. A long term memory circuit that has developed into a cost effective solution is the semiconductor Programmable Read Only Memory (PROM). This memory, along with new integrated circuit analog-to-digital converters and digital-to-analog converters, was successfully used to temperature compensate a crystal oscillator to better than ± 1 ppm from -30°C to $+80^{\circ}\text{C}$.

Introduction

The most commonly used temperature compensation networks for crystal oscillators are analog networks. Two of these, the thermistor varactor networks and the thermistor capacitor networks, are the most widely used. The thermistor varactor type network generates a voltage function which is impressed across a voltage variable capacitance diode (varactor) that is in the frequency determining loop of the crystal oscillator. The voltage changes in such a way that the oscillator remains within a certain specified tolerance of the desired frequency for the temperature range of interest. As we know, this voltage function must be the inverse of the third order Bechman equations for the angle of crystal cut used plus the nonlinear effects of the varactor. Also, it would be desirable to pick up any effects of the individual circuit and crystal used, but this is usually not possible due to increased network complexity. The thermistors and resistors are arranged in networks similar to bias networks that take advantage of the nonlinear resistance versus temperature characteristics of the thermistor to generate the desired voltage function. Usually more than one network is used and the voltages are summed or used in segments switched in or out by transistor switches. In the digital approach a similar type of oscillator is used but with digital techniques to generate the voltage function.

In the thermistor capacitor temperature compensation scheme, thermistors are placed in series with capacitors to effectively change the loop reactance with temperature

and keep the oscillator on frequency. Both of these techniques use thermistors that have historically given the design engineer many headaches. They are very hard to characterize, usually requiring individual testing and sorting to obtain enough accuracy to be used with confidence. The networks used generate systems of equations that require computer solutions in most cases. Then, after the thermistor is in the circuit, it may age or change with time.

There is now a viable solution to this dilemma. Networks have been built using proven digital techniques and no thermistors. The basic block diagram of this network is quite simple although the individual integrated circuits can become quite sophisticated. Indeed, the analog-to-digital (A/D) converters and digital-to-analog (D/A) converters have been the last integrated circuits to mature because of the cross breeding of digital and analog design.

Digital TCXO

The digital technique starts with a thermometer that uses a diode P-N junction to generate a voltage versus temperature function. This is quite linear in the desired compensation range and is $-2\text{mV}/^{\circ}\text{C}$. This voltage is digitized in an (A/D) converter and the parallel digital word is sent to a digital memory. This digital word is used as the address for a new digital word stored in memory or, in other words, the memory is used as a look up table. This new digital word is fed in parallel to a (D/A) converter that generates a voltage to bring the oscillator to the desired frequency. The digital technique generates a segmented approximation to the desired function with accuracy controlled by the number of segments, the number of voltage steps available for compensation and the ability of the circuit to duplicate the original programming. Because this is a segmented approach, some consideration must be given to the number of segments needed. The three criteria needed here are the desired accuracy, the temperature range needed and the steepest slope to be compensated for. For example, if the desired accuracy is one part per million (1 ppm) and the steepest slope to correct is $2\text{ ppm}/^{\circ}\text{C}$, then the correction must be at least every 0.5°C . If the temperature range to be covered is -30°C to $+80^{\circ}\text{C}$ or 110°C , 220 segments are required. The next largest digital word has 256 bits (or an 8 bit byte) which gives us 36 extra bits or

an extra 18°C that can be used to extend the temperature range 9°C on either end. This also has a desirable effect because in the digital approach the compensation stops at the last address so the extended range protects the circuit from falling out of correction too rapidly.

Now for the other half of the job. The oscillator must be pulled back to reference frequency in each segment. Because we don't care what the actual shape of the function looks like, we don't care that any element of the system is linear, only that it is monotonic. We only care that the individual voltage steps are small enough to allow a change of one least significant bit (1 LSB) and still be within desired frequency tolerance and have enough individual words to be able to pull in the farthest excursion within the programming range of the temperature compensation network. For example, if a 1 ppm oscillator were desired and the change of 1 LSB produced a change in frequency of 0.3 ppm, the change would be small enough. Then, if the uncompensated crystal changed ± 12 ppm, at least 72 steps would be required for perfect centering. The next appropriately sized byte would be 7 bits or 128 steps. Extra bits can be used for greater correction, centering or, if available to external logic, to remotely change the oscillator frequency.

It is quite evident by this time that the programming of the memory with 256 digital words for each oscillator would be a prohibitive task if it weren't for automation. This is the great strength of this approach. The programming can be completely automated. The programming effectively tunes out all parts variations and crystal anomalies as long as they repeat in time and temperature.

A programmable calculator with Interface Buss (IB) capability, a counter with IB, and a general digital interface are all that was needed to produce the results to be shown.

Programming the Memory

The automatic system used to gather the data necessary to program the Programmable Read Only Memory (PROM) consists of a CAM programmed environmental chamber, frequency counter, programmable calculator, and PROM programmer. A digital thermometer, digital-to-analog converter and X-Y recorder were also part of the system but were only used to plot frequency versus temperature curves.

The counter is programmed and read over an (IEEE standard 488-1975) interface buss. The calculator can send and receive to the oscillator board over a general digital eight bit parallel interface buss. This buss plugs directly into the PROM socket on the oscillator board.

For the programming cycle, the calculator takes the place electronically of the PROM in the circuit. First, the environmental chamber is driven to beyond the extreme temperature at one end of the working range. Then a four hour CAM is set to take the temperature to beyond the other extreme limit linearly. The PROM programming program is started on the calculator. The temperature of the oscillator is read from its own A/D converter and is a digital number. The calculator creates a memory location with this number as its address, then a digital word is output to the D/A converter on the oscillator board.

The frequency is read by the frequency counter and compared to a reference number, the desired frequency, stored in the calculator. If the frequency is not within prescribed limits, the output word is either incremented or decremented appropriately until the frequency is within limits. When the correct output word is found, its digital value is stored in the memory location of the oscillator's present temperature captured earlier. This cycle is repeated for as many times as available addresses exist in the memory equivalent of the PROM chip. This programming information is saved on magnetic tape. In our set-up the magnetic tape is run through a formatting program that converts this information to paper tape which is in a form acceptable to the PROM programmer. A PROM is blown (or burned) and tested in the oscillator.

Noise

One of the gremlins that plagues a digital oscillator system is noise. This can take two forms; first, there is the susceptibility of the system to noise entering from an external source, and second, the noise generated by the system itself and then retransmitted to the output. The susceptibility of the circuit to noise entering and causing a logic level to change states is a function of the logic family used, but for our purpose is quite small. The second source is quite important. The digital-to-analog converter, by its very nature, is a switched device which in our case also has a preset period during which time it may or may not change states. This output must be filtered because the system interprets any unwanted noise here as modulation impressed upon the primary oscillator loop. The lowest frequency component must be filtered to below system design goals.

A second internal cause of noise is the logic circuits in the system. This noise is generally typified by current spikes on the power supply lines. It is therefore preferable to have the logic circuits on a separate supply buss or connected in ahead of the final voltage regulation and decoupled as thoroughly as possible.

The filtering recommended for logic supply lines by the logic manufacturer is

only adequate to keep one logic device from interfering with another. It is completely inadequate for linear circuitry.

System Timing

The system update frequency determines how often the temperature information is updated, the amount of power drawn by the circuit, and some of the filter parameters. In most applications this can be relatively slow; one or more updates a second should be adequate. For faster systems, the system can be allowed to free convert where the end of one cycle will trigger the start of another. This can be hard wired and requires no external parts, the trade-off being between power and system complexity. If the oscillator were part of a larger system with a microprocessor in control, the system controller could ask for the update when necessary or during processor free time. Most of the circuit could be shut off in between updates as long as the output remained latched.

Oscillator Temperature Stability

The curve at the end of the text is an actual temperature response plotted automatically on an X-Y recorder. Temperature is along the horizontal axis and the output of a frequency counter is sent to a digital-to-analog converter and is plotted along the vertical axis. The apparent unevenness is due to the expanded resolution and the gating of the counter. The extended time over which the plot was taken also has a tendency to pick up stray electrical noise and cause a bad count that is not a function of the oscillator.

The characteristic of segmentation is quite evident. Over a period of decreasing slope the oscillator corrects in each segment giving the impression of a sawtooth wave form. The time constant of the correction voltage can be adjusted to suit the user, keeping in mind that response time could suffer if the time constant is too long. The curve represents our attempt to do better than ± 1 ppm from -30°C to $+80^{\circ}\text{C}$ using digital techniques exclusively.

The resolution of the corrections that can be produced is better than the reproduction of these effects by the oscillator. The limits now become the thermogradient across the circuit and hysteresis effects in the parts used and the crystals themselves. It was evident in working with digital compensation that it is not a new magical cure-all that had the power to throw part tolerances to the wind, but a magnifying glass that amplified the fact that all factors which had to be done right before still had to be designed well, and to gain the potential of this technique, will have to be done better in the future.

Advantages

1. The circuit can be built from readily available parts in CMOS or TTL.
2. The tolerances of critical parts become a "Don't Care" condition, no grading of parts.
3. The process lends itself to mechanized production.
4. Instant On, no warm up time needed.

Future Applications

1. Sharing memory and other hardware with microprocessor systems.
2. Remote tuning of a field oscillator by adding another digital word from a remote location.

References

- 1) A. Mroch and G. Hykes, Collins Radio Group, "A Miniature High Stability TCXO Using Digital Compensation", Proceedings of the 30th Annual Symposium on Frequency Control sponsored by the US Army Electronics Command, p. 292, 1976.
- 2) Buroker, G. E. and Frerking, M. E., Collins Radio Co., "Digitally Compensated TCXO", Proceedings of the 27th Annual Symposium on Frequency Control sponsored by the US Army Electronics Command, p. 191, 1973.
- 3) Sarkar, S. K., Croven Limited, "Explicit Expressions for TCXO Design", Proceedings of the 28th Annual Symposium on Frequency Control sponsored by the US Army Electronics Command, p. 232, 1974.
- 4) Schodowski, S. A., "A New Approach to a High Stability Temperature Compensated Crystal Oscillator", Proceedings of the 24th Annual Symposium on Frequency Control sponsored by the US Army Electronics Command, p. 200, 1970.
- 5) Newell, P. E. and Hinnah, H., Cts Knights, "A Report on Segmented Compensation and Special TCXO's", Proceedings of the 23rd Annual Symposium on Frequency Control sponsored by the US Army Electronics Command, p. 187, 1969.
- 6) Duckett, P. D., Peduto, R. J., and Chizak G. V., "Temperature-Compensated Crystal Oscillators", Proceedings of the 24th Annual Symposium on Frequency Control sponsored by the US Army Electronics Command, p. 191, 1970.
- 7) Boor, S. B., Horton, W. H., Angove, R. B., Systems, Inc., "Passive Temperature Compensation of Quartz Crystals for Oscillator

Applications", Proceedings of the 19th Annual Symposium on Frequency Control sponsored by the US Army Electronics Command, p. 105, 1965.

- 8) Fujii, S. and Uchida, H., Nippon Electric Co., "An Analysis of Frequency Stability for TCXO", Proceedings of the 30th Annual Symposium on Frequency Control sponsored by the US Army Electronics Command, p. 294, 1976.

Block Diagram

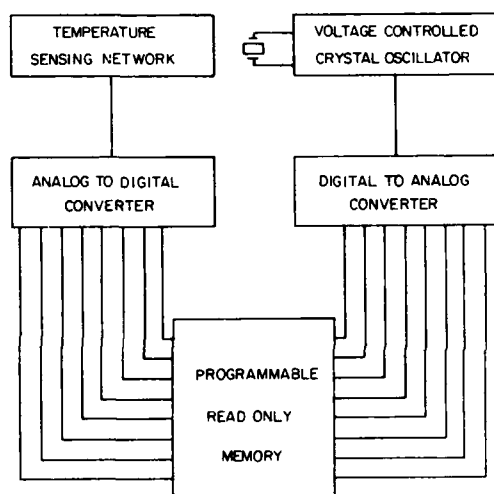


FIGURE 1.

Comparison Chart

	TCXO THERMISTOR VARACTOR	TCXO DIGITAL	OVENS
TEMPERATURE STABILITY -30°C TO +85°C	2.5 PPM	0.5 PPM	0.5 PPM
SIZE	SMALL	MED.	LARGE
CURRENT	LOW	MED.	HIGH
INSTANT ON	YES	YES	NO
CIRCUIT COMPLEXITY	LOW	HIGH IC (MED)	HIGH
COMPUTER CONTROLLED TEMPERATURE COMPENSATION	NO	YES	NO

CALCULATOR BASED AUTOMATIC SYSTEM CONTROLLER AND DATA ACQUISITION SYSTEM

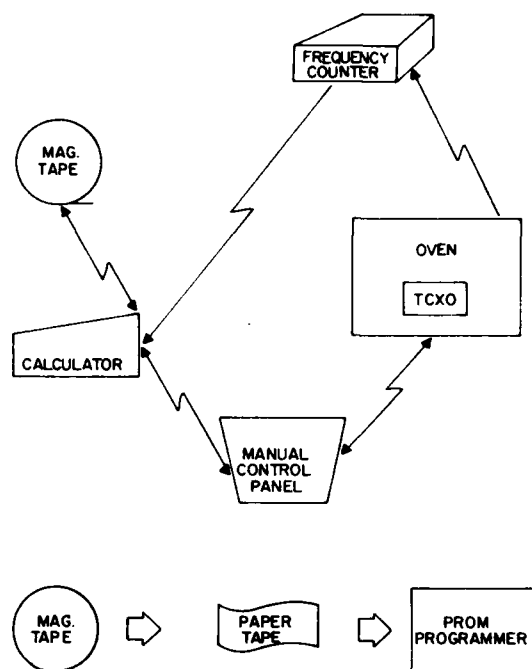


FIGURE 2.

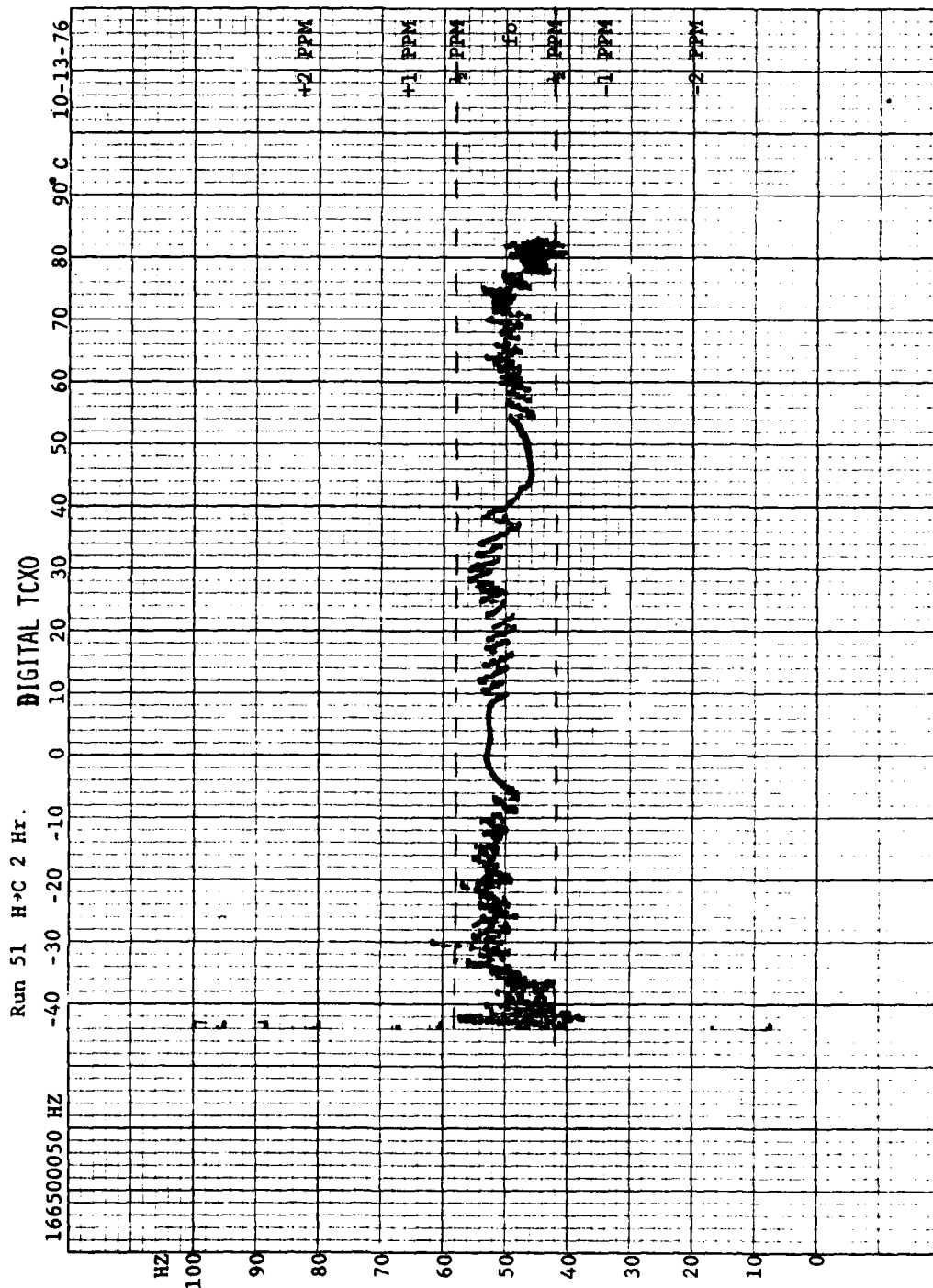


FIGURE 4. X-Y Recording Frequency vs. Temperature

PRECISION OSCILLATORS FLOWN ON THE LES-8/9 SPACECRAFT*

Howard S. Babbitt, III, Group 63
Massachusetts Institute of Technology
Lincoln Laboratory
Lexington, Massachusetts

ABSTRACT

Design, construction, and testing of spaceborne oscillators has been of interest for some time. Although the electrical designs have evolved over a number of years, mechanical construction necessary to survive flight G forces and pyrotechnical shock in addition to actual in-flight aging data are not as readily available. This paper describes the characteristics of four quartz crystal oscillators incorporating Bliley high stability crystals flown on the LES-8/9 spacecraft.

LES-8/9 are a pair of military communication satellites operating at synchronous altitude with multiplied frequencies as high as K-band. Each incorporates redundant master oscillators for time and frequency control. These four master oscillators, built especially for the LES-8/9 program have been measured since launch (March 1976) for absolute frequency drift as well as spectrum quality. To examine long-term drift, a measurement scheme has been employed which to a first order cancels doppler and allows ± 2 resolution of oscillator frequency to within $\pm 3/10^{12}$ in one second. The drift data shows that within one week after launch, the average daily drift for all the master oscillators was between $\pm 7/10^{10}$ /day and $2.7/10^{10}$ /day. One year after launch this had reduced considerably to between $.03/10^{10}$ /day and $.2/10^{10}$ /day. The data presented is believed to include aging only since effects caused by radiation have been minimized by shielding.

INTRODUCTION

Precision ground-based oscillators have been studied for some time with respect to circuit design, packaging, and time and frequency related measurements. With the proliferation of synchronized time systems and high frequency communications links, considerable data has been gathered on very stable frequency sources including precision quartz oscillators.

Recently there has been increased interest in installing stable frequency sources in satellite-based systems. (Ref. 1 and 2) As the satellite functions become more complex and the requirements call for a more stable frequency source, this oscillator must operate under tighter tolerances than its predecessors. Since the number of satellites launched with "precision" oscillators on board is relatively small, in orbit data of actual oscillator systems (Ref. 3-5) is useful for aiding one in assessing what present day capabilities are and what they could be expanded to in the future. This paper addresses four stable quartz crystal oscillators flown on the LES-8/9 spacecraft illustrating how they were built to withstand the launch trauma and the way in which they are being measured in flight.

*This work sponsored by the United States Air Force.

"The views and conclusions contained in this document are those of the contractor and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the United States Government."

LES-8/9 are a pair of military communication satellites which were launched on 15 March 1977 by a Titan 3C launch vehicle. The satellites are presently in a synchronous orbit spaced nearly 90° apart and are used to communicate with up and down links at UHF and K-band with a K-band crosslink (See Fig. 1). They have a five year design life and are powered by a pair of Radioisotope Thermoelectric Generators (RTG's), a source of radiation for all components. Figure 2 shows a general arrangement of LES-9 illustrating the placement of boxes on the main decagon as well as the relative location of the RTG's. With regard to the generation of frequencies from the 5.0 MHz source through multipliers, K-band presented the larger challenge as that represented nearly a X8000 multiplication of the 5.0 MHz standard. Calculating the minimum theoretical increase in phase noise leads to a 78 dB degradation in the signal-to-noise ratio for the oscillator.

OSCILLATOR DESIGN/FABRICATION

Based on requirements for frequency multiplication, frequency settability, and a five year useful lifetime, some minimum specifications can be generated for these oscillators. Several of these can be summarized in Table I. Based upon present day technology, these specifications are achievable with careful oscillator design, crystal selection, and oven design at least insofar as earth-based circuits are concerned.

The circuit selected for the oscillator design was of the Pierce type (Ref. 6-8) and incorporated a Bliley BG61AH-5S high stability crystal as the frequency control element. Figure 3 shows the entire oscillator circuit which was housed in the oven enclosure. The circuit is a rather straight forward implementation of a standard Pierce design with AGC control. It includes two boards, one with the RF amplifiers and gain control element on it and the other with the AGC control circuitry. The AGC control loop bandwidth is approximately 1/2 Hz with an amplitude control of better than .02% for a fixed control voltage on the varactor diode. Over the entire frequency control range the amplitude varies less than 4% and since this represents the maximum change over five years of frequency trimming, this results in an insignificant contribution. The hot carrier diodes, CR4 and CR5, limit the turn-on dissipation to less than 120 μ watts, with the average crystal dissipation near .6 μ watts during normal operation.

In order to take full advantage of the oscillator circuitry and crystal characteristics, an oven was required which would maintain the crystal turning point temperature with a high degree of stability--both long and short term. Some of the specifications for the oven are shown in Table II. Cox and Co., Inc. was contracted to design and build such an oven and enclosure. The specifications were met using a single oven controller with rigid closed-cell freon-filled urethane foam for insulation. This foam also served as a mechanical support for all the inner electronics. The inner electronics included the oscillator circuit boards and crystal oven control board with a regulator for all electronics except the heater voltage. The

circuits were all built on flat donut-shaped boards with the crystal mounted in the middle. The crystal was installed in a cylindrical aluminum housing and that was mechanically attached to an aluminum turning which was the part that was actually temperature controlled. Figure 4 illustrates the packaging of the inner temperature controlled housing. When the entire oscillator systems were put together with the ovens, including RFI filtering in the D.C. lines, buffering, crystal filters, etc. the oscillators performed about as expected. Short term stability was well within specification as was temperature variation and frequency adjustment. Table III summarizes the measurements so far. The unknown fact concerning these oscillators is whether or not they will have sufficiently low long-term drift. We have tacitly assumed that over 5 years the average drift will be less than $1000/10^{10}/5 \times 365$ days = $.55/10^{10}$ /day. This is less than the crystal manufacturer would specify but seemed possible based on the accurate AGC circuitry, the stiff temperature control on the crystal, plus the experience of the GR1115C and other oscillators utilizing this crystal. On the ground it was difficult to determine the actual drift rate as the oscillators needed to be integrated early in the system testing; the satellite assembly was continually being moved, powered down, or measured with respect to power supply loads. Thus, while it was hoped that the above drift rate could be achieved one could never be sure until after launch.

PREPARATION FOR THE LAUNCH AND SPACE ENVIRONMENT

The previous discussion concerning the oscillator/oven combination centered around a basic design which would satisfy the immediate short and long term requirements for the five year mission lifetime. A basic assumption was that the launch trauma as well as long term radiation would not effect these characteristics. Physically, if flown as originally designed, the oscillator/oven would appear rectangular, similar to Fig. 5 with the overall dimensions roughly 6" square by 4" deep. The inner cylindrical oven up to this point has been rigidly supported by the closed cell foam which extends to the outer shell. This rigid structure has considerable mechanical transmissibility from the outer enclosure to the inner oven cavity due to the relative incompressibility of the urethane foam. As a point of reference, Fig. 6 indicates the qualification level vibration which this device must withstand and not suffer any serious consequences. This includes 15.1G RMS overall random vibration for each axis for 3 minutes per axis and 1/2 cycle sine pulse shocks of 775G peak along each axis. Notably, the rigid structure coupled with the 775G peak shock pulse could easily excite resonant modes in the crystal and cause crystal breakage or mount failure.

In view of the stiff requirements for surviving shock and vibration, the oscillator package was analyzed with a shaker table under controlled G-force and waveform shape. A hole was drilled in the middle of one of the covers and an accelerometer was placed on the inner oven enclosure. The accelerometer was then fed back to the measurement system to determine the G-forces seen by this part of the assembly. The initial characterization is shown in Fig. 7. This response to a 1/2g sine-sweep clearly shows unity transmissibility up to 200-300 Hz with a sharp peak of 15g at ~500 Hz. This oscillator/oven assembly with rigid foam all the way to the outer enclosure clearly offers no protection for the crystal or other components in the inner enclosure. Since the mechanical resonance frequency for the crystal blank in its holder is less than 500 Hz, the shock will transmit directly to the crystal and excite this resonance with generally disastrous results. One such prototype oscillator configuration was tested and whereas it withstood the random vibration, the crystal

failed on just the second of 18 possible shocks.

One method to reduce this effect which is popular with those who must minimize stresses of this nature on mechanical parts is shock mounts or isolators. These are small cylindrical or rectangular devices which have threaded studs at each end with rubber connecting them together. As one of the experiments in determining what effect this would have, isolators with about 80 Durometer hardness each were mounted on the shaker table fixture and the box then mounted on them. A sine survey was done as before and the result is shown in Fig. 8. This shows unity transmissibility only up to 25 Hz with a small peak at 30 Hz and every other frequency attenuated. Mechanically this is a very reasonable solution to the problem; thermally it is not acceptable. Since the spacecraft body and the rails to which boxes must be attached are temperature controlled, the thermally defined surface would be only indirectly coupled through the rubber isolators. This thermal resistance is high and therefore the oscillator enclosure could soar in temperature and easily exceed the maximum practical operating temperature of 50°C.

If, therefore, due to thermal considerations, the enclosure must be attached directly to the spacecraft rails, then the next level of mechanical isolation must be inside. Since the crystal itself is mounted rather securely in its own aluminum housing and this is directly connected to the inner oven aluminum piece, it was rather difficult to isolate just the crystal from shocks. The solution which was ultimately used for flight was to cut approximately 1/4" of the rigid foam away from the outer enclosure and install soft foam strips in the space. Figure 9 illustrates this clearly and shows a picture of a flight package with one cover removed. The soft foam is ECCOFOAM FS, an isocyanate open-cell foam sheet stock which was cut in strips wide enough and thick enough to maintain the proper thermal resistance for correct temperature control. A sine survey was run on this assembly and the result is shown in Fig. 10. This frequency response was done with a 1G input level as opposed to Figs. 7 and 8 which were with 1/2G inputs. The results show unity transmissibility up to approximately 70 Hz with a soft peak at 110 Hz. Other frequencies were attenuated, except for a lonesome narrow peak at ~800 Hz. Since there is considerable attenuation between 200 Hz and 700 Hz, the area of crystal mechanical resonance, the soft foam strips seem to have solved the problem. At this point, a measurement of the total response to shock pulses was in order. A 1/2 cycle sine pulse of 775G's and .2 MS duration was applied to the structure. An accelerometer again recorded vibration levels and the frequency components measured. Figure 11 shows the result with the input spectrum superimposed on the data. It clearly indicates at least a factor of fifteen attenuation to most of the energy, essentially guaranteeing that the circuits would survive. Subsequent tests with qualification level vibration and shock on a flight-like oscillator/oven showed no problems whatsoever, with pre- and post-vibration frequency differences of less than $1/10^{11}$.

Magnetic Effects

It is well known that magnetism has an effect on absolute frequency. In LES-8/9 a large number of magnets were in close proximity to these oscillators. Although an exact calibration of the sensitivity of the circuit to magnetic fields was not done, a test magnet which produced 2 to 3 gauss at the oscillator was used to determine if there was sensitivity at all. In the case of these oscillators, there was a disturbance of the absolute frequency on the order of parts in 10^8 . A sufficient amount of shielding was added until this effect was considerably smaller than parts in 10^{11} .

This shielding turned out to be .025 inches of Conetic AA μ -metal. It was installed just on the inside of the outer enclosure and is visible in Fig. 9. With this additional metal on the box, the actual flight weight was nearly 4 pounds.

Shielding

With the crystal in its own aluminum can attached to the inner oven, there is a considerable amount of aluminum surrounding the crystal. If to this you add the thickness of the cover and the μ -metal, the minimum amount of metal which can be considered as crystal shielding is greater than .46" of AL and .025" of μ -metal. Synchronous attitude spacecraft do not run in the Van Allen Belt so that radiation is not a hazard, but the on board RTG's do provide radiation which could be harmful. Based on these thicknesses of materials, calculations have shown that the electron energy received at the crystal is confined to less than 5 MEV electrons with a total 5 year dose of $< 1.6 \times 10^6$ electrons/cm² or less than 2×10^{-3} RAD in 5 years. There is no significant proton, neutron, or gamma radiation. It is for these reasons that it is felt the radiation effects are sufficiently minimal as to be non-existent in causing frequency changes in the oscillator even over a 5 year lifetime.

MEASUREMENT IN-FLIGHT

After the successful launch of LES-8/9 in March, 1976, all the measurements, components, care in assembly, etc. were put to the test of actually performing to specification for the required 5 years. Short term stability and spectrum content of the K-band oscillator were measured by inference through BER (bit-error rate) tests and direct downlink spectrum analysis. The long term drift presented more of a problem for two reasons. First, to maintain all the full back-up modes the oscillator outputs could not be swapped for fear of glitching a clock line and disturbing some active storage. This meant that the oscillator which provided the downlink frequencies was the only one which could be directly received at an earth terminal. Figure 12 illustrates the general interconnections. Second, the oscillator which is seen at the earth is corrupted by doppler. This doppler is as much as $.5/10^6$ for the LES-8/9 orbit and when compared to resolutions desired of better than $1/10^6$ of the oscillator frequency any small doppler uncertainties would completely mask out the quantity we are desiring to measure. Typically, an orbit determination is made by observation and/or ranging and a computer calculates doppler as a function of time and day. This has all sorts of uncertainty built into it and in fact even in the most sophisticated doppler computing schemes .1% uncertainty is not uncommon. Thus, some sort of doppler cancelling scheme is needed which is not based on computer predictions but is a closed loop measurement system in real time. The doppler cancelling method used on LES-8/9 to measure absolute satellite frequency will be described first with the problem of measuring the "unseen" oscillator after that.

Theory of Measurement of Absolute Oscillator Error

The basic premise which initiated this measurement scheme is as follows: If the satellite is transmitting a downlink frequency, the ground station knows what that frequency should be. With the naive assumption that the on-board oscillator is perfect, the frequency error is only due to doppler. Therefore, if an uplink is sent to the satellite and frequency is precorrected for this doppler, the signal should enter the satellite's receiver at the exact center frequency assuming the receiver is driven by the same oscillator as the transmitter. The question arises, how do oscillator errors

affect this measurement? Certainly they change the situation somewhat, but one would think if the errors are small enough, the same basic arguments hold. That is in fact the case, and as will be shown, the satellite receiver now would receive a frequency which is in error by approximately twice the oscillator error. Thus, a measurement of the center frequency error could be a measure of oscillator error. LES-8/9 has the ability to encode baseband I & Q samples and send these down directly on the downlink so the receiver channel frequency error can be examined to verify this. With this introduction, we will analyze the simple scheme of Fig. 13.

The symbols as described on the figure include

c = Relative oscillator error

d = Relative doppler error

f_{xy} : $x = R$ for "Received"

T for "Transmitted"

$y = S$ for "In the Satellite"

E for "On the Earth"

f_{REF} = Satellite Reference Frequency

With the assumption that all RF Frequency outputs are derived by multiplication from the same basic oscillator, we have

$$f_{TS} = N f_{REF} \text{ and} \quad (1)$$

$$f_{RS}^o = M f_{REF} \quad (2)$$

We will first proceed assuming that $c = 0$. For that case

$$f_{RE} = N f_{REF}^o (1+d) \quad (3)$$

where $f_{REF}^o = 5.0 \times 10^6$ exactly.

Since $c = 0$ and we were expecting $N f_{REF}^o$, then the percentage doppler error must be

$$d = \frac{f_{RE} - N f_{REF}^o}{N f_{REF}^o} \quad (4)$$

Transmitting on the uplink and making our doppler correction, we have

$$f_{TE} = (1-d)M f_{REF}^o, \text{ or} \quad (5)$$

$$f_{TE} = \left(2 - \frac{f_{RE}}{N f_{REF}^o}\right) M f_{REF}^o \quad (6)$$

Now, if we compute what the satellite receive frequency is, we find

$$f_{RS} = (1+d) f_{TE}, \text{ or substituting (5)} \quad (7)$$

$$f_{RS} = (1-d^2) M f_{REF}^o \quad (8)$$

or doppler has been cancelled to a first order with only second order terms in the equation. If Eq. (6) is used as the algorithm for computing f_{TE} even if $c \neq 0$, the following results. In all cases

$$f_{RS} = f_{TE} (1+d). \quad (9)$$

Substituting from (6)

$$f_{RS} = M f_{REF}^{\circ} \left(2 - \frac{f_{RE}}{f_{REF}^{\circ}} \right) (1+d). \quad \text{But} \quad (10)$$

$$f_{RE} = N f_{REF}^{\circ} (1+\epsilon) (1+d). \quad (11)$$

$$f_{RS} = M f_{REF}^{\circ} [2 - (1+\epsilon) (1+d)] (1+d), \quad (12)$$

or simplifying

$$f_{RS} = M f_{REF}^{\circ} [1 - \epsilon - 2\epsilon d - d^2 (1+\epsilon)] \quad (13)$$

Now, on the satellite, the center frequency of the receiver is

$$f_{RS}^{\circ} = M f_{REF}^{\circ} (1+\epsilon). \quad (14)$$

Forming the difference between the actual received frequency and the true center frequency of this receiver,

$$f_{RS} - f_{RS}^{\circ} = -M f_{REF}^{\circ} [2\epsilon + 2\epsilon d + d^2 (1+\epsilon)]. \quad (15)$$

$$\text{Since } d \leq .5/10^6,$$

$$d^2 \leq .25/10^{12}. \quad (16)$$

By similar arguments, we can also say

$$2\epsilon(1+d) \sim 2\epsilon. \quad (17)$$

Thus, (15) simplifies to the desired result.

$$f_{RS} - f_{RS}^{\circ} = -M f_{REF}^{\circ} (2\epsilon). \quad (18)$$

Equation (18) tells us that if we listen to the satellite downlink, assume all the error is doppler only, adjust our earth station uplink to cancel this doppler to a first order, then the frequency received by the satellite will be in error by twice the absolute frequency error of that uplink receiver. To measure that frequency error, the satellite has the capability to encode in-phase (I) and quadrature (Q) outputs and SSB modulate these bits onto the K-band downlink. These bits are then demodulated at the earth station for processing. In practice, Eq. (6) is implemented through the use of mixing and phase-locked loops with dividers. The downlink demodulation is done with a PSK demodulator followed by digital filtering on the I and Q lines. The final output has a resolution of 1/4 Hz (since we have I and Q samples) or since this represents twice the error of the satellite, the ultimate resolution is 1/8 of a Hz of satellite error. If the K-band uplink is on the order of 40 GHz, then this represents a resolution of $3/10^{12}$ for a one second averaging time. The actual hardware performing the counting integrates this error over approximately 13 seconds to reduce the jitter and ease the computation required.

Measurement of the "Unseen" Oscillator

The master oscillator which is not directly visible and hence not measurable by the above method necessarily must be measured indirectly. In the satellite the two master oscillators are compared at 1.6 MHz with an up-down counter. This up-down counter is strobed once a

telemetry format and the register data set to the ground station. The inference of the unseen oscillator is done by making two observations of this counter spaced a known difference in time apart. This leads to a difference frequency which is equal to the count difference divided by the product of 1.6×10^6 and the time differential in seconds. An elapsed time of approximately three hours leads to an average uncertainty of less than $5/10^{11}$. Better resolution can be had by observing this count as a function of time and noting exact transition times.

Results of Long Term Drift Measurement

Drift measurement of all four precision oscillators on both satellites were made commencing approximately three days after launch using the methods described above. Although the absolute error was not allowed to exceed $\pm 5/10^{10}$ for any oscillator, the result of the necessary frequency trimming has been added back in to get overall drift curves. Data has been taken for over one year and so can give insight into long term drift characteristics of some types of spaceborn oscillators. The data are shown in Figs. 14-17. The fastest changing oscillator--LES-8, OSC #1--changed over $120/10^{10}$ in the year, whereas the smallest overall change was observed on LES-9, OSC #1 where it drifted only $50/10^{10}$ in one year. If these are maximum changes for a year, and we have no reason to believe they would get worse, then there will be plenty of margin to cover 5 years as each oscillator has at least $1000/10^{10}$ trimming range. All oscillators were set to be on frequency at the time of launch. If the drift curves are projected back to the day of launch, none of them seem to have suffered the $10/10^{10}$ incremental change which would have been expected due to the 1G field change.

Scanning all four oscillator's drift curves LES-8 OSC #2 stands out as being unusual. For the majority of the time the oscillator is extremely stable. On two occasions, however, large frequency excursions were experienced in a relatively short time. The cause of this is unknown, but could be due to some component relaxing mechanically (i.e., a piston tuning capacitor). The other three oscillators have relatively smooth aging curves which appear to be reducing somewhat in slope even after a year in orbit.

Figures 18-21 are the derivative functions of Figs. 14-17. It is simply a two-point slope calculation of the absolute drift data. The drift rates appear to be exponentially decaying with time constants on the order of 30 to 60 days. LES-8 OSC #1, settled out to the apparent final drift rate in under 100 days. That number is smaller than $+.3/10^{10}$ /day. LES-8 OSC #2 even prelaunch had a small drift rate; if the impulsive functions are discounted, the average drift rate has been well under $1/10^{11}$ /day. In fact from eight days after launch until 88 days after launch, the oscillator was not trimmed and accrued a net increase of less than $.2/10^{10}$ for an overall drift rate for that period of much better than $1/10^{12}$ /day. That oscillator has only been trimmed during two periods: from day 88 to day 110 and on day 339. The variation in the curve at other intervals is the uncertainty in the "unseen" oscillator measurement. LES-9 OSC #1 and OSC #2 have been tracking one another remarkably well, although OSC #1 has a slightly lower drift rate. Both curves indicate that the drift rates are still decreasing in with OSC #2 presently between $.5/10^{11}$ and $1/10^{11}$ /day and OSC #1 better than $3/10^{12}$ /day.

Conclusions

The results presented in this paper cover two basic areas: packaging and measurement of precision satellite-based oscillators. The oscillator packaging described here is somewhat different from the more usual oven design in at least two ways. First, it is not a "thermos bottle" or Dewar flask container, but simply an inner aluminum oven surrounded by closed cell foam with soft foam for vibration isolation. Second, it is a one stage oven which maintains $\pm 0.001^\circ\text{C}$ internal temperature stability while the external environment changes 60°C . This is contrasted with the more popular double oven approach used on some recent satellite oscillators.

The measurement of these oscillators with regard to frequency drift is a peculiar problem requiring careful attention to error introduced, particularly doppler. The closed-loop doppler cancelling scheme used to measure the LES-8/9 master oscillators has allowed absolute frequency measurements to be made of these oscillators with uncertainties on the order of parts in 10^{12} even though they are more than 20,000 miles away. The results show that for the 5.0 MHz oscillators flown on LES-8/9 the total yearly drift was less than $120/10^{10}$ for any oscillator with drift rates at the end of one year as low as $3/10^{12}$ /day.

REFERENCES

1. Proceedings of the 30th Annual Symposium on Frequency Control, U. S. Army Electronics Command, Fort Monmouth, NJ, (1976). Copies available from Electronics Industries Association, 2001 Eye Street, NW, Washington, D.C. 20006.
2. Proceedings of the 8th PTTI Applications and Planning Meeting, December, 1976.
3. L. J. Rueger, "One-way Dissemination from low Altitude Satellites" Proceedings of the 25th Annual Symposium on Frequency Control, U. S. Army Electronics Command, Fort Monmouth, NJ, pp. 179-185 (1971). National Technical Information Service Accession NR. AD 746211.
4. R. Easton, C. Bartholomew, J. Bowman, M. Block, "A Second Satellite Oscillator Experiment" Proceedings of the 24th Annual Symposium on Frequency Control, U.S. Army Electronics Command, Fort Monmouth, NJ. pp. 339-344 (1970). Copies available from Electronics Industries Association, 2001 Eye Street, NW, Washington, D.C. 20006.
5. T. B. McCaskill and J. A. Buisson, "NTS-1 (Timation III) Quartz and Rubidium Oscillator Frequency Stability Results", Proceedings of the 29th Symposium on Frequency Control, U.S. Army Electronics Command, Fort Monmouth, NJ. pp. 425-435 (1975). Copies available from Electronics Industries Association, 2001 Eye Street, NW, Washington, D.C. 20006.
6. Chirlian, P. M., "Electronics Circuits: Physical Principles, Analysis, and Design", pp. 716-739. McGraw-Hill Book Company, New York, (1971).
7. Stratemeyer, H. P., "The Stability of Standard-Frequency Oscillators", The General Radio Experimenter, Vol. 38 #6, June 1964, pp. 1-16.
8. Pustarfi, H. S., "An Improved 5 MHz Reference Oscillator for Time and Frequency Standard Applications", IEEE Transactions on Instrumentation and Measurement, Vol. IM-15, #4, December 1966, pp. 196-202.

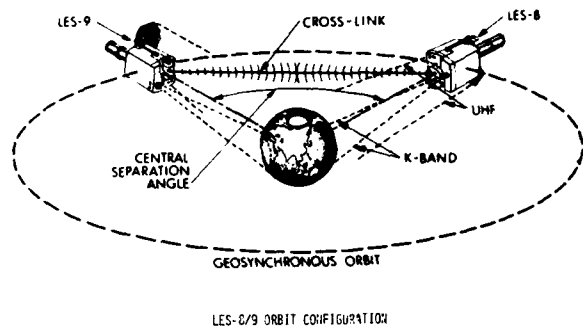
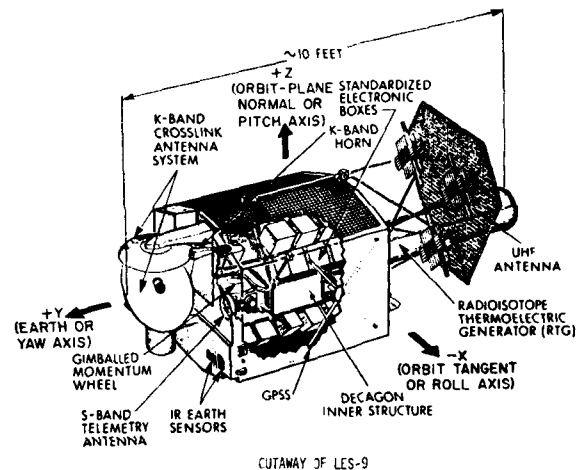


FIGURE 1



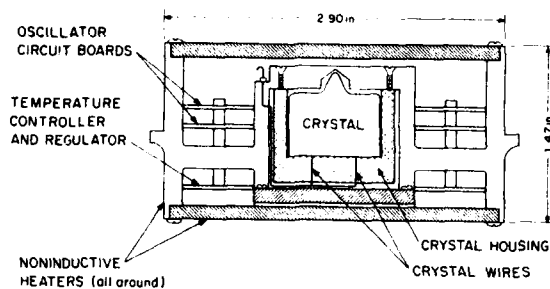
CUTAWAY OF LES-9

FIGURE 2



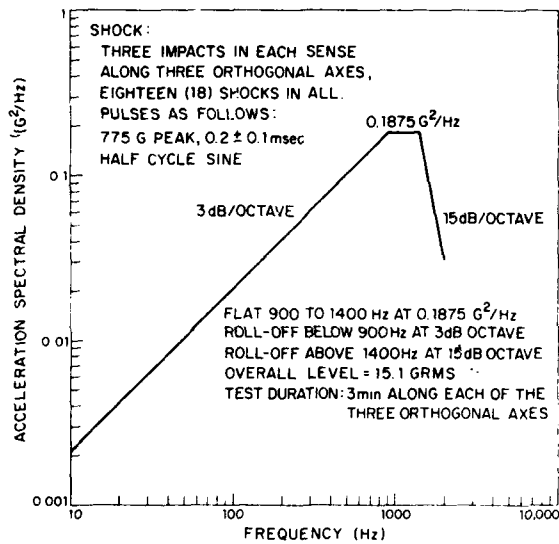
GENERAL SCHEMATIC OF BASIC OSCILLATOR

FIGURE 3



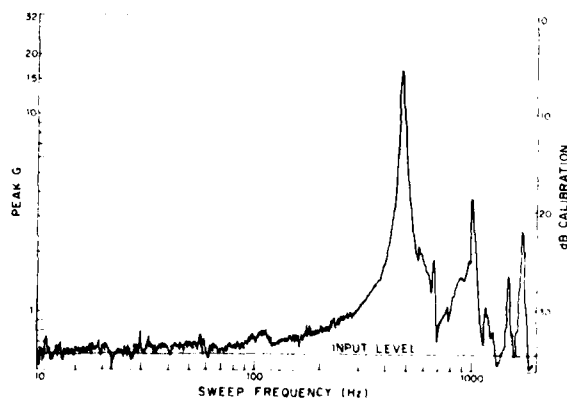
INNER ALUMINUM OVEN DETAILS

FIGURE 4



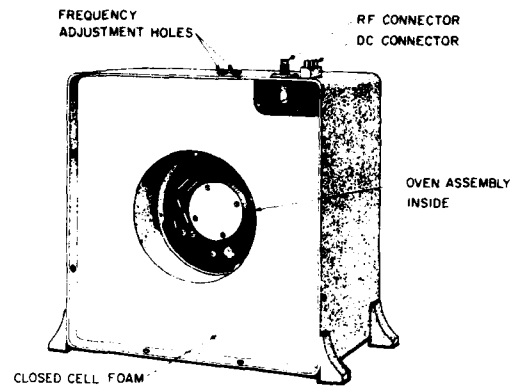
SATELLITE RANDOM VIBRATION QUALIFICATION TEST

FIGURE 6



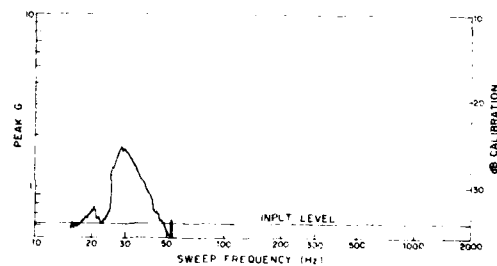
SINE SURVEY OF ORIGINAL OSCILLATOR/OVEN MOUNTED DIRECTLY TO VIBRATION TABLE

FIGURE 7



OSCILLATOR AND OVEN WITH NO VIBRATION PROTECTION WITH ELECTRONICS COVER REMOVED

FIGURE 5



SINE SURVEY OF ORIGINAL OSCILLATOR/OVEN WITH SHOCK MOUNTS INSTALLED BETWEEN BOX AND VIBRATION TABLE

FIGURE 8

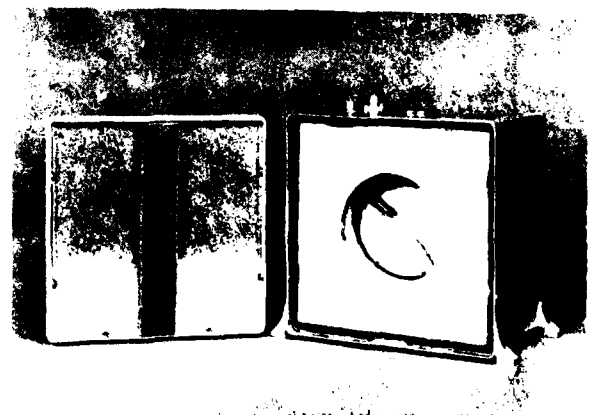
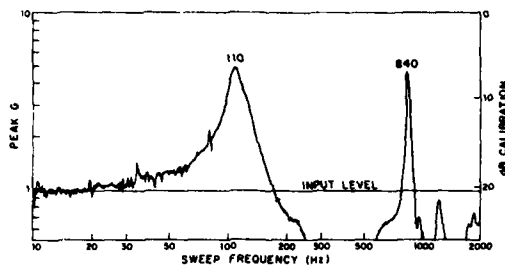
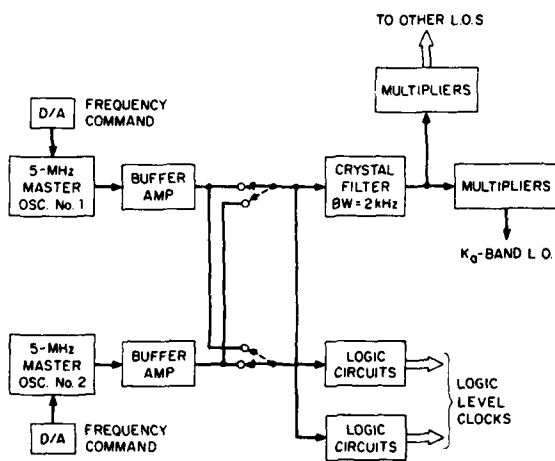


FIGURE 9



SINE SURVEY OF MODIFIED OSCILLATOR/OVEN
WITH SOFT FOAM STRIPS SUPPORTING RIGID FOAM INNER STRUCTURE

FIGURE 10



SIMPLIFIED SCHEMATIC OF THE MASTER OSCILLATORS
AND THEIR INTERCONNECTION TO RF AND LOGIC CIRCUITS

FIGURE 12

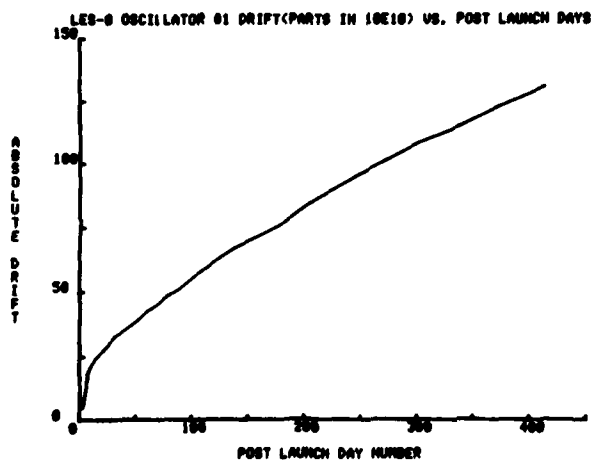
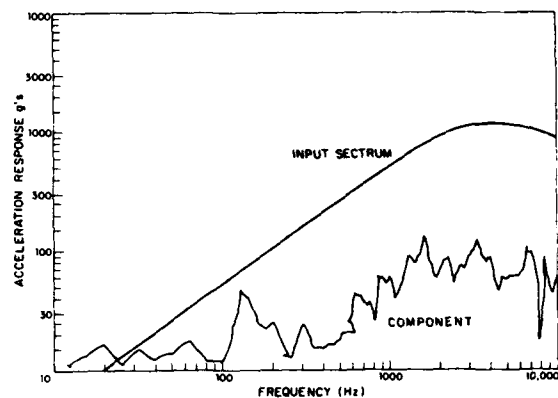
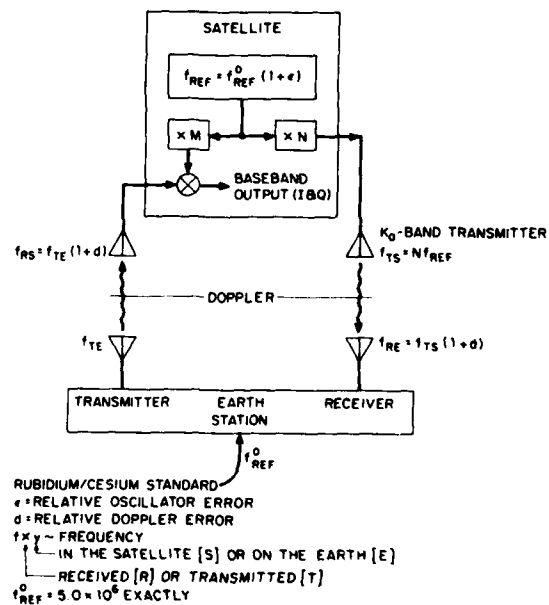


FIGURE 14



FREQUENCY RESPONSE DUE TO SHOCK PULSE OF
MODIFIED OSCILLATOR/OVEN
WITH SOFT FOAM STRIPS SUPPORTING RIGID FOAM INNER STRUCTURE

FIGURE 11



GENERAL SCHEMATIC FOR ANALYSIS OF
OSCILLATOR DRIFT MEASUREMENTS

FIGURE 13

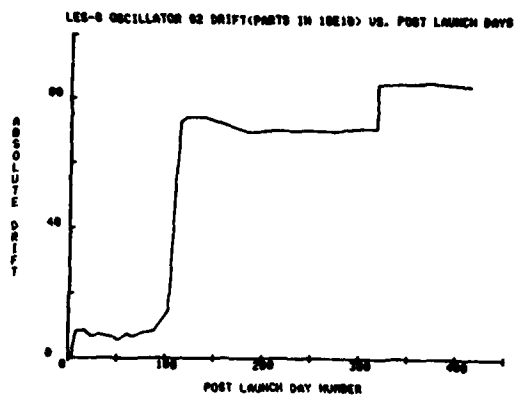


FIGURE 15

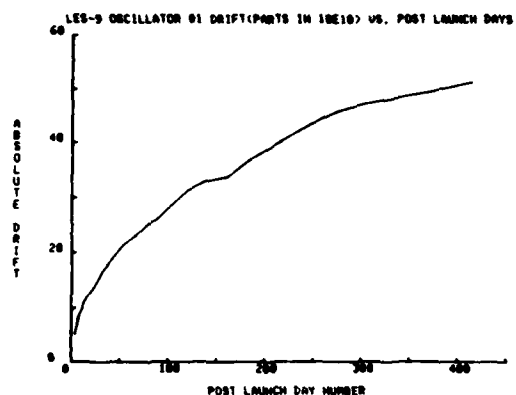


FIGURE 16

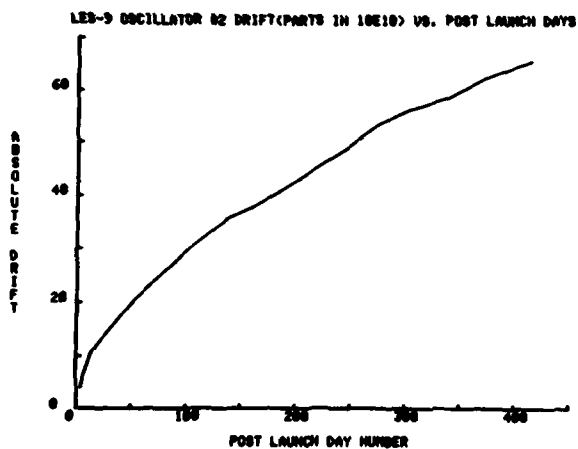


FIGURE 17

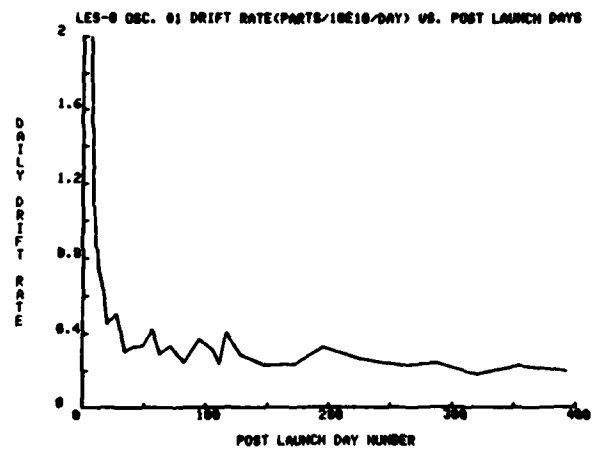


FIGURE 18

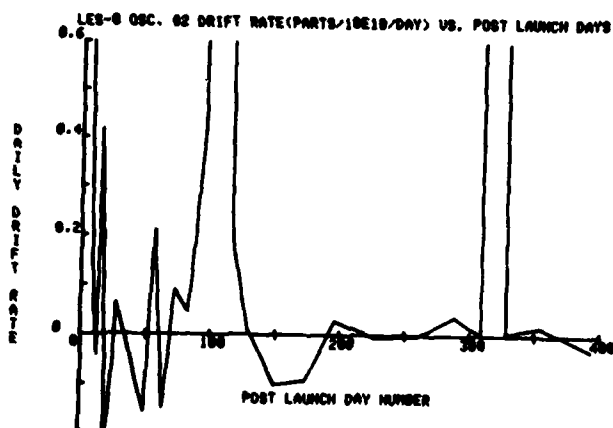


FIGURE 19

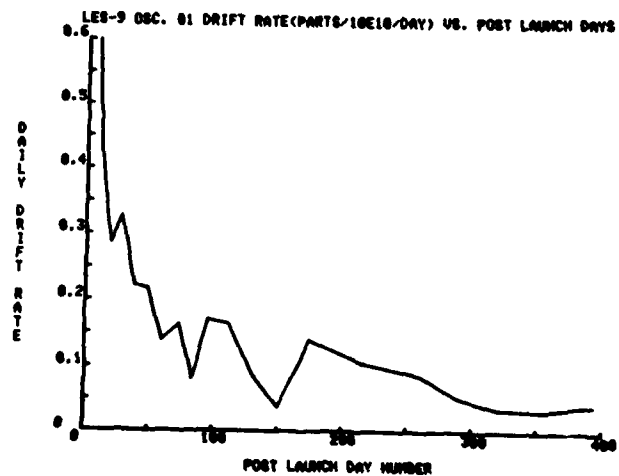


FIGURE 20

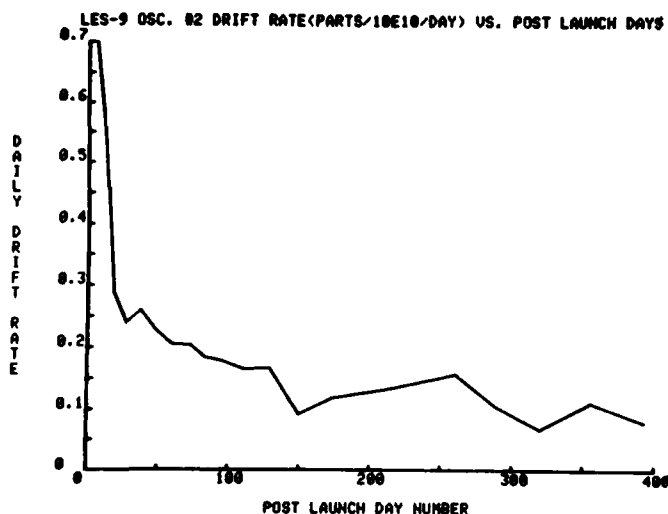


FIGURE 21

SUMMARY OF SELECTED MASTER OSCILLATOR REQUIREMENTS FOR LES-8/9

Frequency = 5.0 MHz

$$\left. \frac{\Delta f}{f} \right|_{\text{RMS}} \ll 1.2/10^8/100 \mu\text{sec}$$

$\sim 3/10^9/100 \mu\text{sec}$ design goal

Drift Rate < $1/10^{10}$ /day

Absolute Frequency Adjustment < $1/10^{10}$ /step

Total Adjustment Range: Sufficient to maintain absolute frequency accuracy for 5 years

TABLE I

PRELAUNCH OSCILLATOR CHARACTERISTICS

$$\left. \frac{\Delta f}{f} \right|_{\text{RMS}} < 6/10^9/100 \mu\text{sec}$$

OSCILLATOR OVEN SPECIFICATIONS

Control Temperature:	71°C to 79°C
Temperature Stability:	±0.001°C from set point
Ambient Variation:	-20°C to +40°C
Temperature Accuracy: (5 year life)	±0.050°C

TABLE II

Remote Frequency Adjustment:	10 V, 12 bit D/A applied to the varactor diode
Frequency Adjustment Steps:	Between $7/10^{12}$ and $2/10^{10}$ depending upon center frequency
Total Commandable Frequency Range:	> $1000/10^{10}$
System Temperature Sensitivity:	< $1/10^{11}/\text{C}^0$ of ambient change

TABLE III

A FAST WARMUP OSCILLATOR FOR THE GPS RECEIVER

John Ho and Martin Bloch
Frequency Electronics, Inc.
New Hyde Park, New York

Summary

A 5.115 MHz oscillator is described which was designed for use in the GPS receiver system. The oscillator features fast warmup, high stability, low power, dual RF outputs, and $\pm 2 \times 10^{-7}$ modulation range. The device uses a third overtone FC or SC cut crystal with low temperature coefficient. A booster oven in addition to the conventional dual oven is used for fast repeatable warmup characteristics.

Introduction

This paper deals specifically with the mechanical and thermal design aspects of the oscillator and test data is presented which describes the warmup characteristics, aging data and short term stability. A complete analysis of thermal paths due to conduction, radiation, and convection as well as thermistor placement and oven loop parameters is required to yield the desired performance. Although the power consumption of the oscillator was not a fundamental design constraint, reduction of this parameter was considered in order to minimize power supply capacity and size.

Electrical Design

The basic oscillator design focuses upon minimizing the warmup characteristics of the oscillator to enable the GPS receiver to acquire lock within 5 minutes. Additionally the stability of the oscillator must be low enough to insure a reasonably short processing time of position information. The requirements are as follows:

1. Input power. 10 watts for 3 minutes, 1 watt after 15 minutes.
2. Warmup time. 2×10^{-8} 5 minutes after turn on, 1×10^{-9} 1 hour after turn on.
3. Aging. 2×10^{-9} per day, 30 minutes after turn on and 2×10^{-11} per day average.
4. Output power. + 3 dbm per output.
5. Size. 3.20x1.60x1.60 inches.
6. Weight. 8 oz.

Figure 1 is a block diagram of the oscillator. An internally generated reference voltage is fed to user equipment. The modulation input enables the frequency to be varied over a range of $\pm 2 \times 10^{-7}$. Failure of the inner or outer oven is indicated by utilization of a temperature monitor window consisting of thermistors and an amplitude comparison circuit. Variations of greater than 1°C are considered a failure. The oven booster uses the raw D.C. input to provide a quick surge of power

for rapid warmup. The thermal design of the oscillator which is within the inner oven in conjunction with the booster oven input is selected to give a critically damped response in the frequency vs. time domain. This interface must be carefully controlled to avoid frequency overshoot which would result in a prohibitive sinusoidal frequency excursion.

The oscillator circuit is a modified Pierce crystal controlled oscillator. A third overtone FC or SC crystal is used. The oscillator and crystal are enclosed within the ovens which in turn are mounted within the Dewar flask. The proportional oven controllers maintain a temperature variation of the inner oven of less than 1.0 millidegrees after stabilization. The oven controllers combine a thermistor sensing bridge and operational amplifier to generate the required heater currents.

The voltage regulator and oven control circuitry are external to the dual assembly and provide regulation and filtering of the D.C. input to minimize induced noise.

Mechanical Characteristics

Figure 2 is a cross section view of the oscillator. The basic mechanical package consists of a ruggedized Dewar flask, shock mounted inside a metal can. The configuration is specifically designed to pass severe shock and vibration testing. The flask has a reinforced tip at its open end to prevent cracking during shock and random vibration. In addition, an asbestos lining between the walls of the flask is held in place by dimples in the outer wall which are located at the center of gravity of the package. The flask is held in place with rubber shock mounts which are placed equidistant from its center of gravity to equalize bending moments.

In order to minimize size and weight, hybrid circuitry is used for the oscillator and buffer amplifier. The oven controls and voltage regulator are monolithic integrated circuits.

Figure 3 is the outline of the crystal oscillator. The RF interface is a SMA connector and all logic, D.C., and modulation voltage are fed to the unit through FRI filters.

Oscillator Performance

This section describes the test results obtained on several of the GPS oscillators and discusses the crystal testing that is undertaken in order to insure proper oscillator performance.

Figure 4 is the short term stability measured over a sampling time of 100 microseconds to 10 seconds. It is seen that the stability above 1 second approaches 2×10^{-12} and continues at this value until aging effects become dominant at several hundred seconds.

Figure 5 shows the phase noise measured from 10 Hertz to 20 KHz from the carrier. Two plots are shown which compare oscillators with and without an output filter. The inherent noise floor is -144 dbc which is reduced to -169 dbc at frequencies greater than 8 KHz from the carrier by the filter. The filter has a 3 db bandwidth of 100 Hz.

Figure 6 shows the warmup characteristic at +25°C for the GPS oscillator. It is seen that at 5 minutes after turn on the nominal frequency is reached and at 6.8 minutes a maximum excursion of 5×10^{-9} is reached. This performance is well within the GPS requirements for the receiver clock. Due to the high inrush current an overshoot condition is obtained with a maximum negative excursion of 2×10^{-7} .

Figure 7 is the aging data for the same oscillator. Data is presented for a 20 week period and it is seen that the aging slope levels off to a value of 2×10^{-11} .

Crystal Characterization

In order to insure that the oscillator meets the stringent warmup and aging requirements each resonator is tested in a fixture which is contained within a temperature stabilized chamber. Figure 8 is a block diagram of the crystal measurement system. The oven temperature is automatically slewed and approximately 2,000 data points are taken at 0.05°C increments. A network analyzer automatically records frequency and resistance on a punched tape. This tape is fed to a calculator and a XY recorder where a plot is obtained. Figure 9 is a typical graph showing a smooth temperature vs. frequency characteristic with minimal retrace. In addition the resistance holds constant over the temperature range of 0 to +70°C. This is an ideal crystal for use in the described oscillator. Figure 10 shows a faulty crystal where the resistance is changing rapidly at the turning point. This type of behavior would result in oscillator performance that is unacceptable.

Figure 11 shows an oscillator characteristic where anomalous behavior occurs at +5°C. This is acceptable since the inner oven temperature for this device is set at 51°C, the crystal turning point. Figure 12 illustrates unacceptable oscillator performance where anomalies are noted at +32°C and in the area of the turning point at +52°C through +60°C.

It should be emphasized that frequency vs. temperature anomalies occur over very narrow temperature ranges and it is impera-

tive that a suitable measuring system be utilized in order to detect these problems.

Conclusion

A 5.115 MHz oscillator has been described which is optimized for frequency retrace characteristics and quick warmup capabilities. In addition low aging and power consumption are obtained. The role of fast warmup precision oscillator in modern receiving systems is extremely important when considering the vast numbers of satellite user equipment that will be implemented in the next decade. The use of oscillators such as described in this paper will greatly enhance the operational capabilities of these receivers.

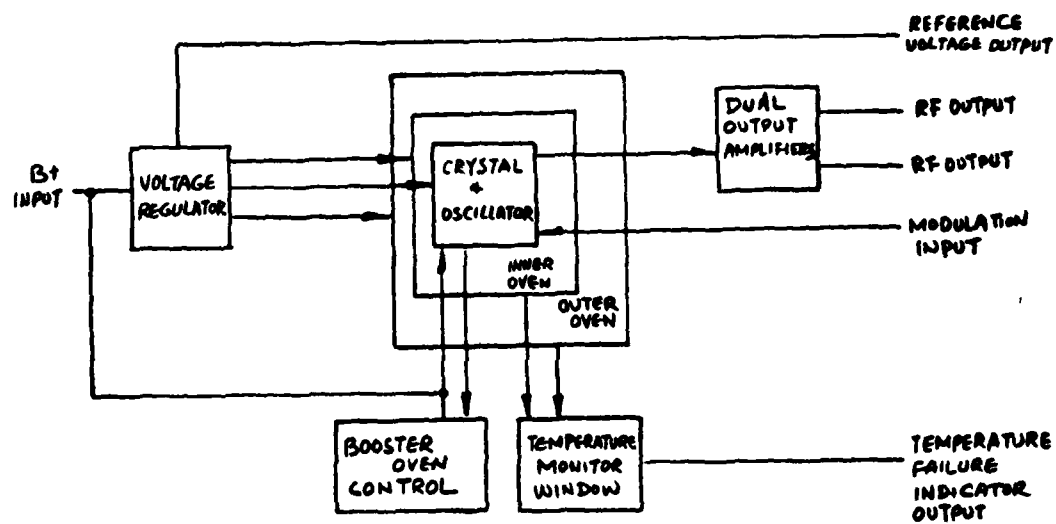


FIGURE 1. BLOCK DIAGRAM, MODEL FE-2140A

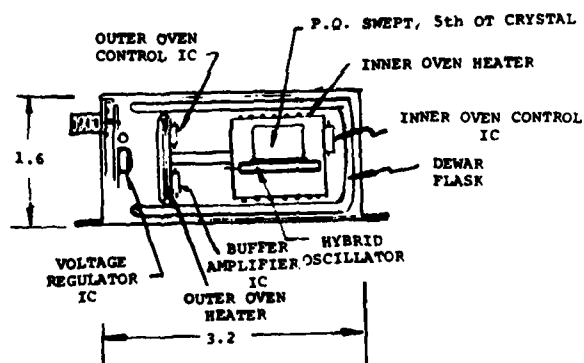


FIGURE 2. CUT-AWAY VIEW, MODEL FE-2140A

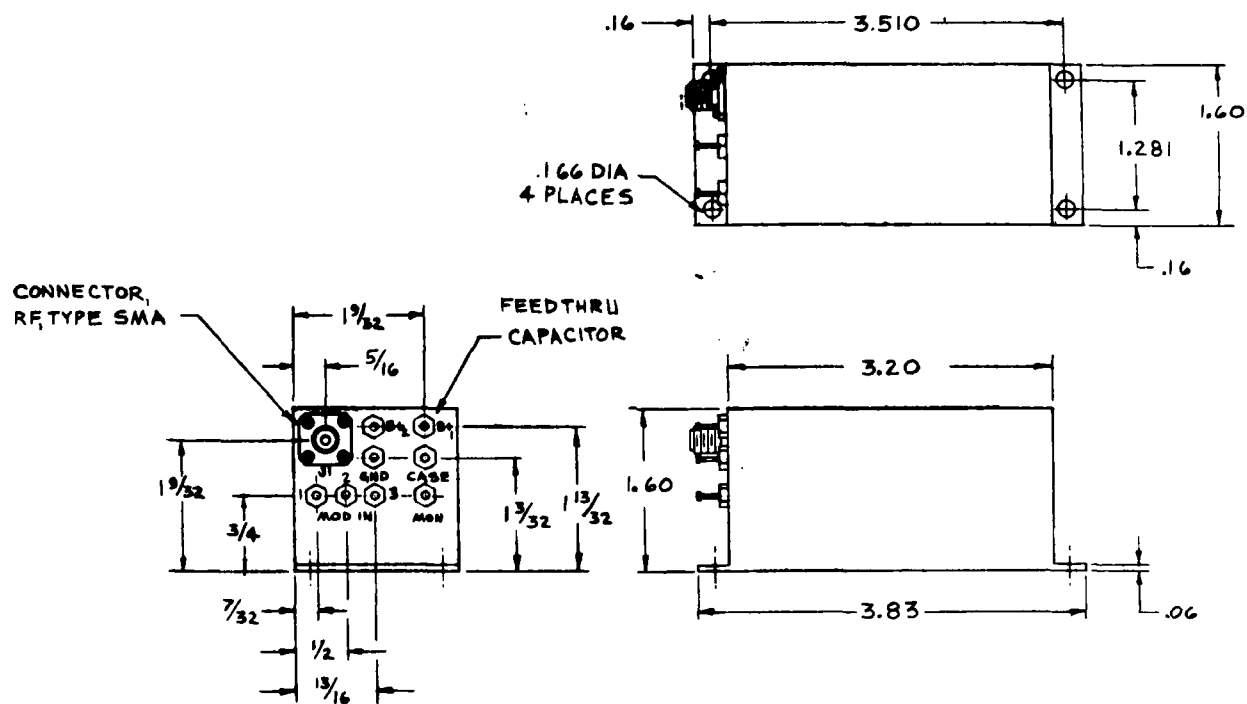


FIGURE 3. OUTLINE DIMENSIONS,
MODEL FE-2140A

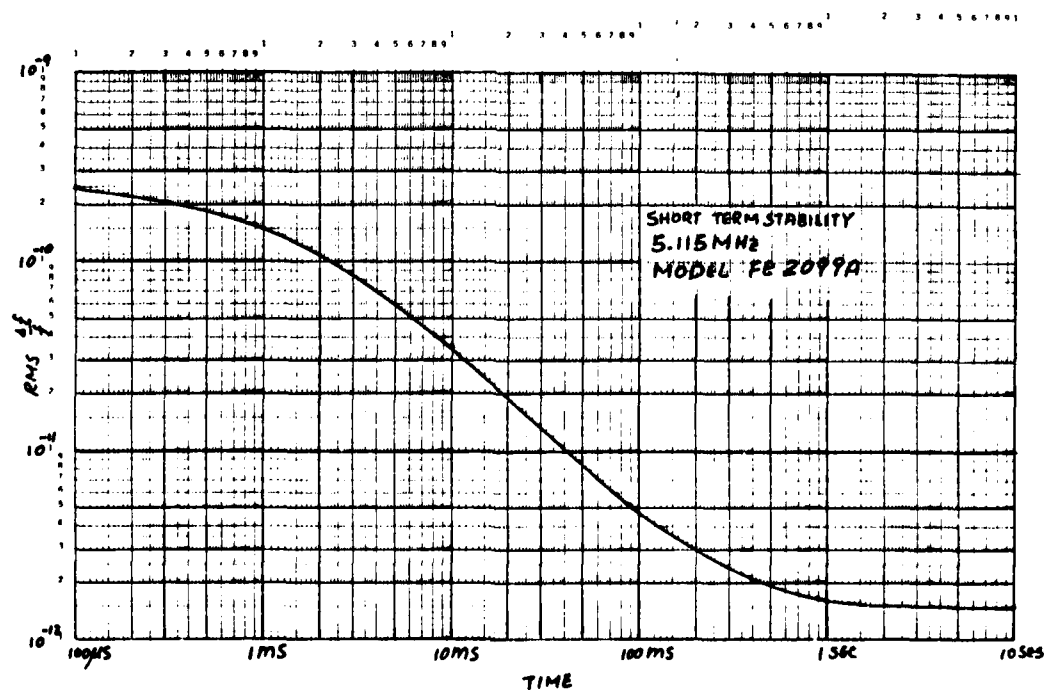


FIGURE 4. SHORT-TERM STABILITY

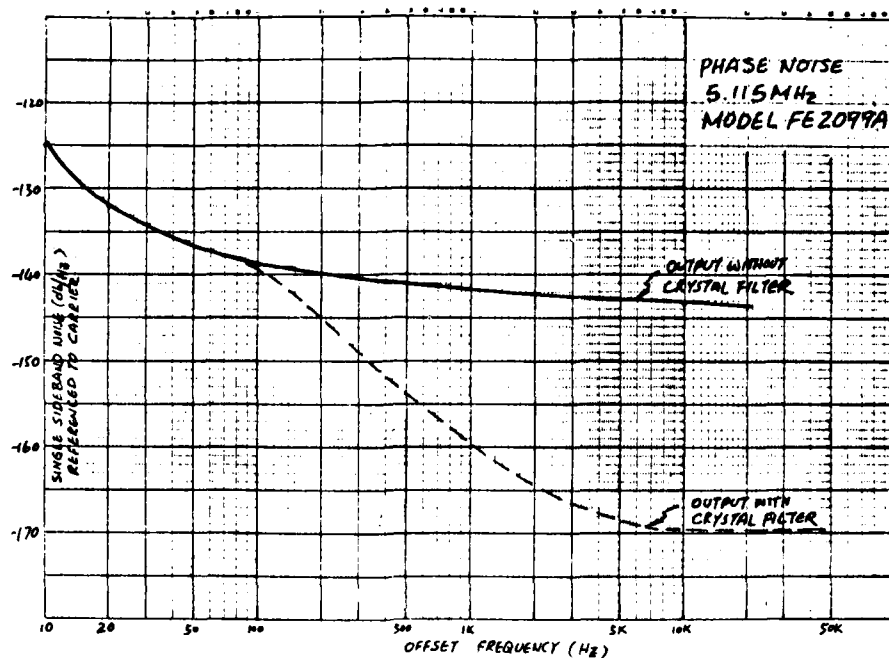


FIGURE 5. PHASE NOISE

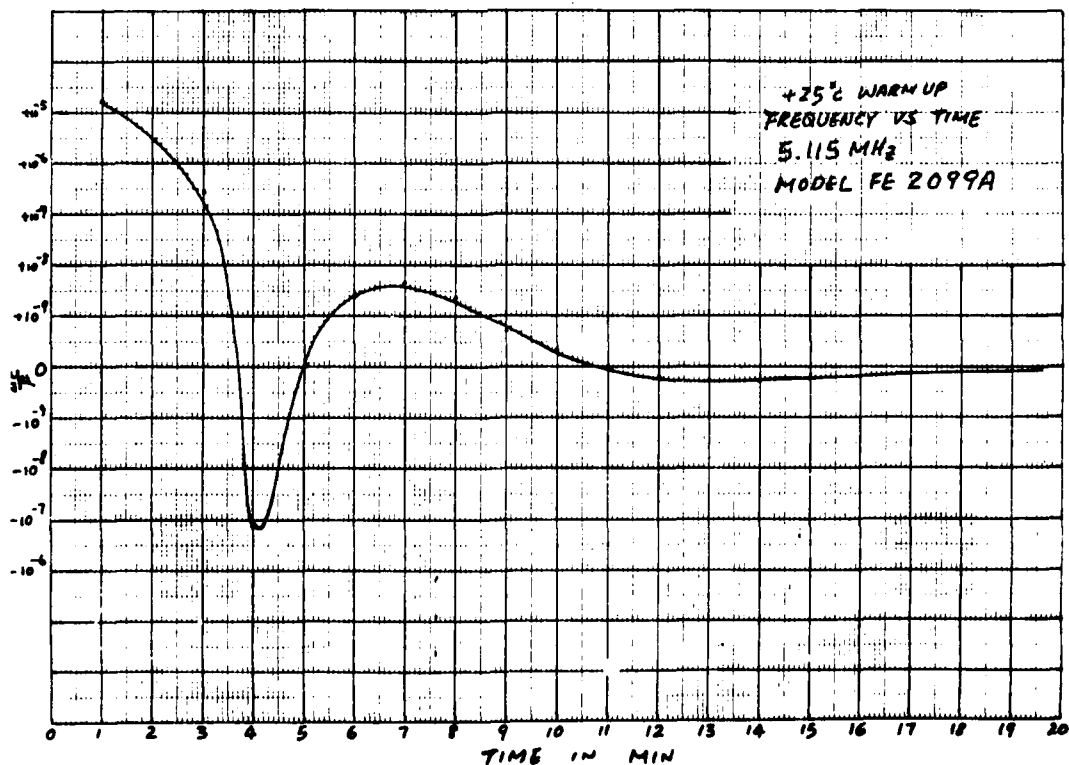


FIGURE 6. WARMUP CURVE

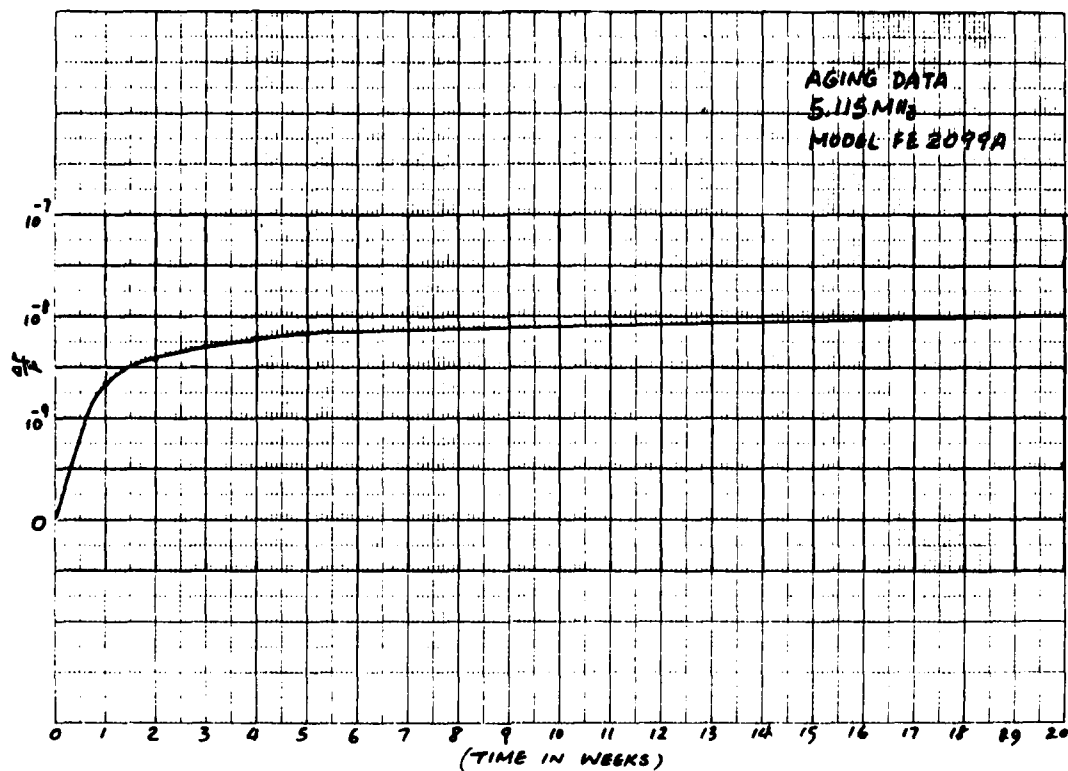


FIGURE 7. AGING CURVE

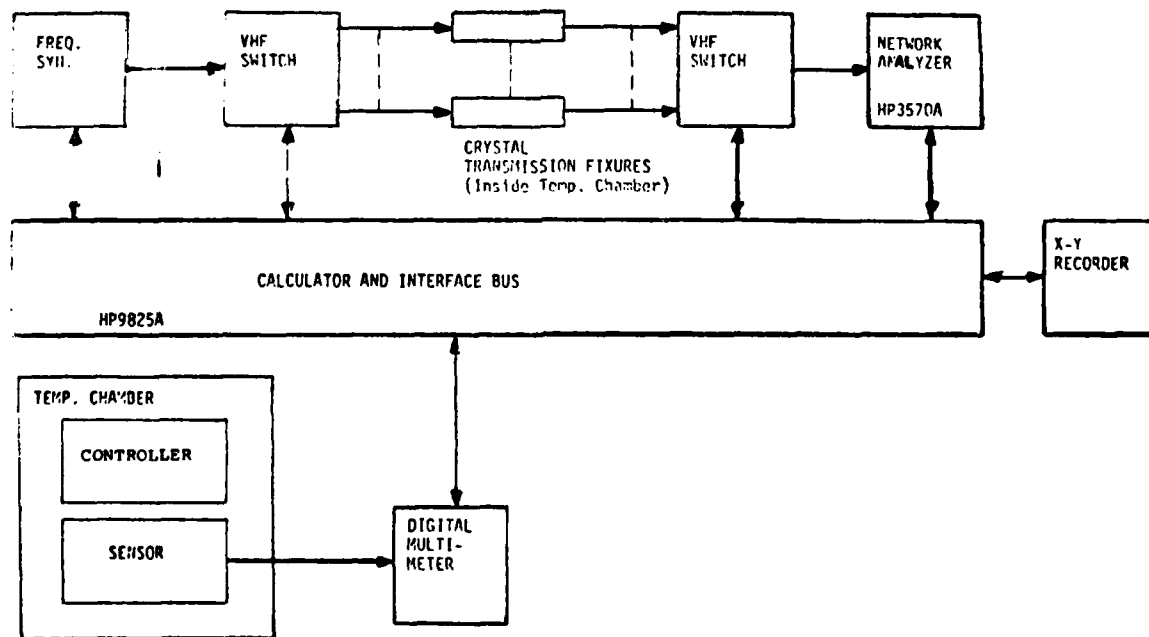


FIGURE 8. CRYSTAL MEASUREMENT SYSTEM

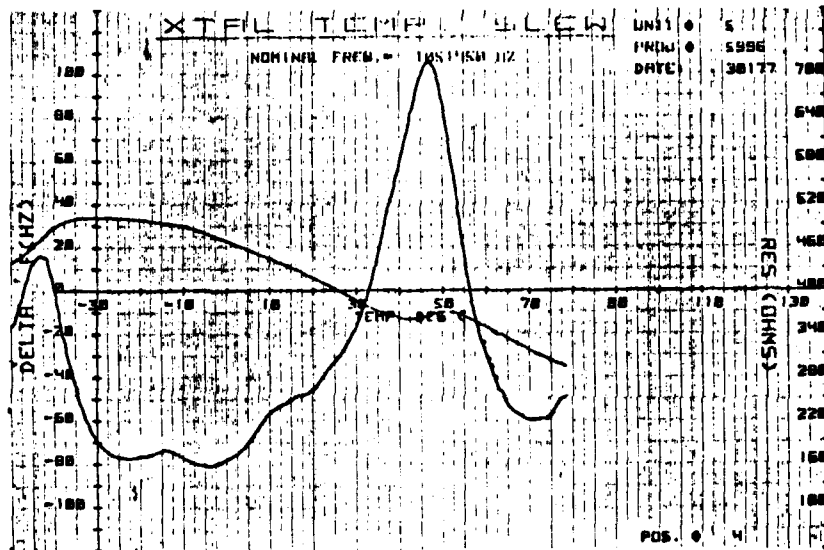


FIGURE 9. CRYSTAL TEMPERATURE
vs. FREQUENCY

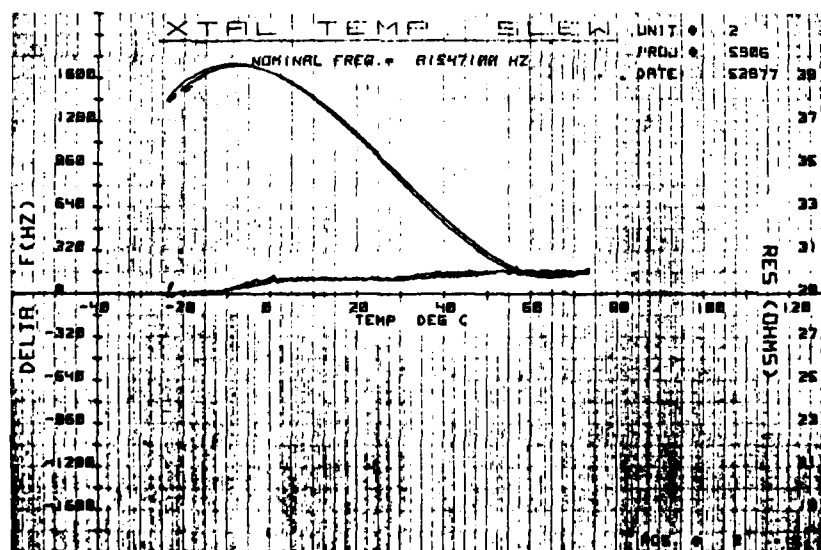


FIGURE 10. CRYSTAL TEMPERATURE
vs. FREQUENCY

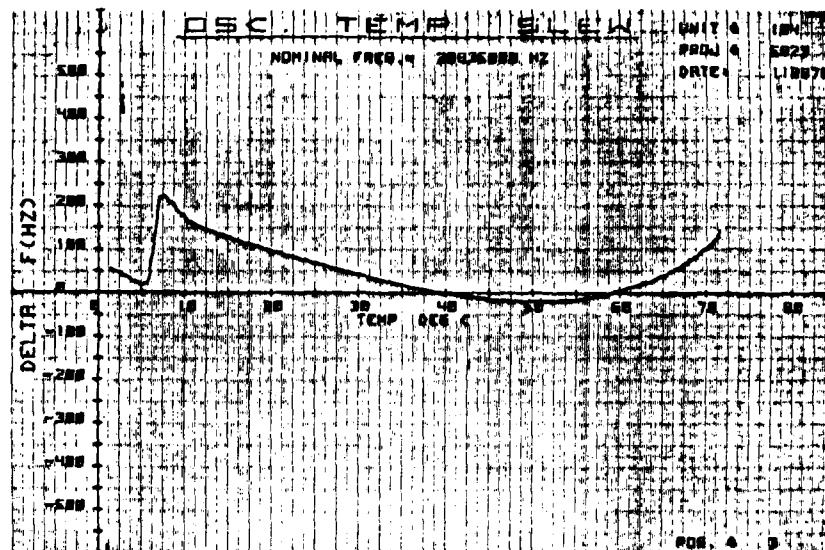


FIGURE 11. OSCILLATOR TEMPERATURE
VS. FREQUENCY

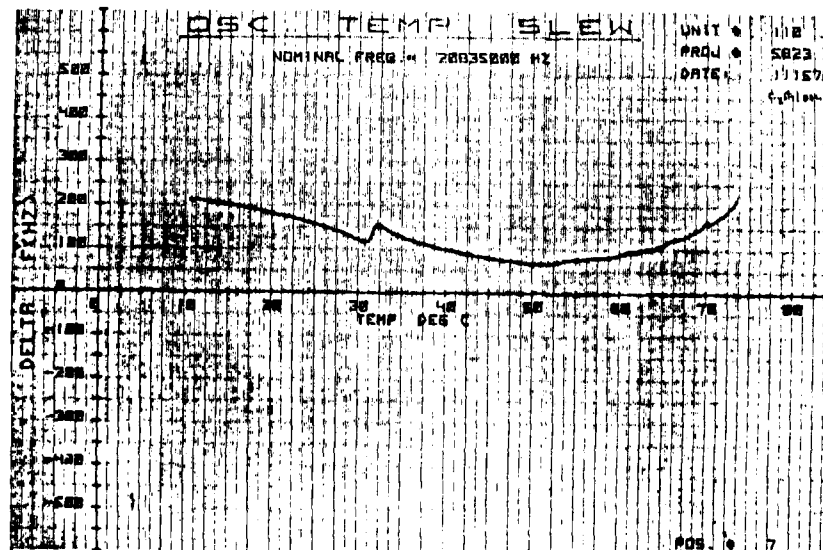


FIGURE 12. OSCILLATOR TEMPERATURE
VS. FREQUENCY

SYNCHRONIZATION METHODS FOR FREQUENCY- AND TIME-DIVISION-MULTIPLEX NETWORKS

H.L. Hartmann
Lehrstuhl für Nachrichtensysteme
Technische Universität Braunschweig (FRG)

Summary

This contribution examines both Directed and Mutual Synchronization for Time-Division-Multiplex (TDM) and Frequency-Division-Multiplex (FDM) communication networks. A systematic analysis of mutually synchronized oscillator networks leads to additional conditions with regard to the attainment of a sufficient line-jitter attenuation and of a tolerable phase offset in the network.

As a guiding figure of merit with respect to long- and short-term inaccuracy of synchronized networks, the time error $\Delta t = \langle \Delta t \rangle + \delta t$ between different network outputs is considered. The long-term inaccuracy is described, then, in terms of the normalized and generally time-dependent ensemble average $\langle \Delta t \rangle / t$ and the short-term inaccuracy in terms of a time-independent normalized standard deviation $\langle (\delta t)^2 \rangle^{1/2} / t$.

In the case of synchronized FDM and TDM networks it is possible to keep the above defined long-term inaccuracy minimal in comparison to that of a free-running atomic frequency-standard. Caution is, however, advised in the evaluation of short-term stability, because in this case, either the filtered phase-jitter of the transmission paths or the amplified frequency noise of the oscillators tends to dominate in the national or international network areas, respectively.

Basic Methods of Network Synchronization

All methods of synchronization are intended to attain the regeneration of a frequency or time-base, generally after signal transmission over radio- or cable-systems including channel noise and link interruptions. Desirable are those methods of regeneration which facilitate a nearly complete reestablishment of the reference signal, so that the interfering influences of the transmission systems become negligible, and the quality of the reference source remains unchanged. The application of multi-channel carrier frequency transmission systems with frequencies of from 12 to 200 Mc/s as well as the establishment of public or non-public communication systems for digital signals with transmission rates of from 600 bit/s up to 500 Mbit/s pose the question for manufacturers of the systems and administrative companies as to how a reliable, uncomplicated, and exact distribution of frequency and time-bases can be maintained and of how the detailed functions of the systems can be supervised. Figures 1 and 2 demonstrate the relations between the required frequency accuracy and the magnitude of frequency-deviation (FD) on the one hand, and between frequency accuracy

and time-deviation (TD), on the other, as a function of varied frequencies (f) and of slip intervals (t), respectively. Table 1 offers a survey of the already-known methods of synchronization of communication networks. The following basic methods can be singled out:

Directed Synchronization by means of Single or Changeable Major Timing Sources and Mutual Synchronization according to the so-called Single- or Double-Ended Principle.

Conventional Directed Synchronization is analogous to the classical Master-Slave principle, in which the timing supply of a network is originated by one central timing source. In the advanced methods of directed synchronization timing sources are changeable, that is to say the reception of timing for the entire network can be transferred to a remote alternate timing source. This helps to secure uninterrupted availability of timing signals. Generally, the switching from one reference timing source to another results in the rearrangement of the phase relations in the entire network. In either case, the individual dependent timing sources must not interpret a reception interference as a total failure of the major timing source. Instead, they must accept the timing information through an alternate connection. The patent application for this method was filed in 1960 by G.P. Darwin and R.C. Prim¹.

In the case of Mutual Synchronization, a variety of time sources in one intermeshed network are mutually tuned to a common frequency. The intermeshing of the timing sources insures a high degree of security against interruptions in the signal transmission paths. In this way, each source is capable of exerting only minor influences on the resulting common network frequency. A total failure of one source amounts to the isolation of only one node in the network. The basic idea that an estimated mean of incoming interfered phases can be processed in place of a single time reference was originated by J. P. Runyon in 1960².

One must not forget that, in the design of a system which operates on such a phase-averaging principle, the phase propagation times of the oscillator network can lead to an effect on the resulting common network frequency which requires attention. It has, however, been shown that, either by single-ended, digitally-controlled reference phase prediction (digitally controlled Single Ended-System (SES)) or by double-ended propagation delay time measurement with compensation (Double Ended System (DES)), correct operation of the system can be

guaranteed^{3,4,5,6}. Both of these procedures possess the property of being able to tune geographically-separated communication centers to a common network frequency. The availability and the short- and long-term stability of the frequency is, above all, dependent upon the chosen synchronization method and, in addition upon the design of the control loops which are utilized. Instead of treating all of the many possible special solutions to these problems, we shall discuss the consequences of a general solution which is approximately applicable to all of them. The basic configuration for the regeneration of timing is shown in figure 3. It corresponds to a generalized phase-locked loop which has been extended to more than one input. W refers to the Weighting of the timing, ACC refers to an Accumulation with respect to time, VCO refers to the Voltage Control of the Oscillator and ISC to the digital Control for the Initial State which includes Supervision, SUV.

Results of a system-theory of network synchronization

The common differential equation for a controlled oscillator or network node (i), controlled as shown in figure 3, can be written, under the simplifying assumption of linear and time-invariant system components and with the normalized phase $p(t) = \varphi(t)/2\pi$, as

$$\dot{f}_i(t) = f_{i0}(t) + h(t, \tau_h) * K \sum_{j=1}^N a_{ij} [p_j(t - \tau_{ij}) - p_i(t) - \Delta p_{ij}(t)]. \quad (1)$$

In this equation, $K \cdot a_{ij} = K_0 \cdot K_0 \cdot a_{ij}$ represents the open loop gain of one input (j) referenced to the output (i) in the steady state; $h(t, \tau_h)$ represents the impulse response of the accumulator with the time constant τ_h ; and a_{ij} is the weighting factor of the input j at network node i. K_0 and K_0 are measures for the dynamic or steady state slope of the control characteristic of the phase discriminator or of the oscillator, respectively. The dimension of K is (rad/s)/rad=Hz. $\Delta p_{ij}(t)$ represents a measured or predictive reference phase²³, and $f_{i0}(t)$ represents the freerunning frequency of the oscillator (i).

The system of equations (1) generally comprises $i=1$ to N equations for the computation of the unknown $f_i(t)$ and $p_j(t)$. It has been demonstrated in several publications under rather general assumptions with respect to weighting factors and values for the reference phase, $p_{ij}(t)$, that the solution for the entire network which this system represents assumes a common frequency $f_i(t) = f$ and a stationary phase distribution $p_i(t) - p_j(t) = p_i - p_j$. Particularly, it has been possible to develop realizations of equation (1), which insures linear operation of the entire network without excluding the application of digitally working phase-locked loops^{3,4,5,6,10}.

For communication engineers two aspects of equation (1) are, with respect to time, of special interest:

1. Long-term behavior of the network for $K^{-1} < t < \infty$:

In this case, equation (1) assumes the simplified form

$$\dot{f} = f_{i0} + K \cdot \sum_{j=1}^N a_{ij} [(p_j - p_i) - f \tau_{ij} - \Delta p_{ij}] \quad (2)$$

Network frequency f , then, approximately follows the relation

$$f = \langle \hat{f}_{i0} \rangle / (1 + K \langle \Delta \tau_{ij} \rangle)$$

if $\langle \hat{f}_{i0} \rangle$ is assumed to be the estimated ensemble mean of f_{i0} and $\langle \Delta \tau_{ij} \rangle$ that of all delay-time variations $\Delta \tau_{ij}$ in the network²³.

2. Short-term behavior of the network for $0 \leq t \leq K^{-1}$:

In this case, an algebraic form can only be obtained by the Laplace-Transformation $L\{p(t)\} = P(s)$ of equation (1) into the domain of complex baseband frequency $s = (j\Omega - 2\pi f) + j(\Omega - 2\pi f) = \omega_i + j\omega_b$. Equation (1), then, assumes the form

$$s P_i(s) = F_{i0}(s) + H(s, \tau_h) K \sum_{j=1}^N a_{ij} [P_j(s) e^{-s\tau_{ij}} - P_i(s) - \Delta P_{ij}(s)]. \quad (3)$$

The whole system of equations, after a few computations, can, through a combination of all output-phases $P_i(s)$ and because of $s\tau_{ij} \approx 0$ be written in a matrix notation

$$P(s) = [E - G(s)]^{-1} \cdot G(s) \left[\frac{F_{i0}}{K \cdot H(s)} + \Delta P(s) \right] \quad (4)$$

where $G(s)$ is the matrix of the weighting factors $K \cdot H(s) \cdot a_{ij}$. This equation illustrates the relation of the output phases $P(s)$ of an oscillator network to its inner phase sources, $F_{i0}(s)$ and $\Delta P(s)$. The matrix product $[E - G(s)]^{-1} \cdot G(s)$ describes the closed-loop gain of the entire network. One can see that, in addition to the poles of $H^*(s)$ and $G(s)$, those of the matrix $[E - G(s)]^{-1}$ become part of the stability condition. Gersho and Karafin have demonstrated⁷, that stability of the entire system is insured, if

$$\text{Det} [E - G(s)] \neq 0 \quad \text{for } s > 0 \quad (5)$$

and

$$\lim_{s \rightarrow 0} \text{Det} [E - G(s)] \xrightarrow{s^1} 0 \quad (6)$$

i.e., if the system, for $s=0$, shows only a pole of first order. The validity of equation (5) results in stability against gradual rising oscillations; that of equation (6) in stability against frequency drift. In these cases it is suitable to speak of short- and long-term stability. This is insured if: a) for every sum of line-elements of (s) (i.e., the maximum closed loop gain of every generalized control loop), according to figures 3 and 4 in connection with 0, the equation

$$|g_i(s)| = \left| \sum_{j=1}^N g_{ij}(s) \right| = \left| \frac{K \cdot H(s) \sum_{j=1}^N a_{ij}}{s \cdot K \cdot H(s) \sum_{j=1}^N a_{ij}} \right| < 1 \quad (7)$$

is valid; and b) if this value, for $s \rightarrow 0$,

approaches 1 with no higher degree than s^1 . But, this leads to the condition of peak-jitter free control-loops with an inherent phase error. Fig. 5 demonstrates the situation of the poles and the measured dynamic behavior of a network of four variously intermeshed oscillators as a function of the product Kt_h . The stability of mutually-synchronized networks with $H(s) = (1 + s t_h)^{-1}$ is secured, if $Kt_h \leq 0,5$ for SES or $\leq 0,25$ for DES, respectively.

Observations concerning figures of merit for various methods of synchronization in communication networks.

Figure 10 demonstrates nominal short- and long-term inaccuracy of various frequency sources. Short-term frequency variations up to 100s are generally expressed in terms of the Two-Sample Allan Variance $\sigma_y(\tau)$ and long-term variations as the peak coefficient of $\Delta f/f$ for $t > 100$ s. Both values are based on estimations denoted by $(\hat{\cdot})$. In order to keep the frequency offset in FDM networks or the time offset in TDM networks small, the application of Cesium atomic-frequency standards emerges as a highly exact and available frequency source. For a discussion of this principle, see reference¹¹.

The following examples of methods of synchronization are intended to show what steps can be taken to help insure that the above-mentioned quality characteristics will be retained. As a reference network, we chose a hierarchically organized international network with an attached national network geared to present-day and future communication networks, in this case that in the Federal Republic of Germany (figure 6). In addition to the networks on an international level (1) one must also take into account networks on the national level (2 to M), according to their size (1 to M-1). On the international level, references to the propagation-delay-time variations are based on the conclusions obtained through the operation of the satellite "Symphonie"¹². On the national level, estimated means of propagation-delay-time variations are based on those existing now in the Federal Republic of Germany. Possible configurations for the two basic methods of synchronization are presented in figures 7.1 and 7.2. Their respective merits can be described in terms of the following characteristics:

1. The ensemble average of the fractional time- or of the frequency-deviation ($\Delta t/t$ or $\Delta f/f$ respectively), between different network node outputs derived from the phase drift of the networks, as a figure of merit for long-term-inaccuracy.

2. Variance of the relative time difference between different network node outputs, derived from the variance of the phase variations of the network, as a figure of merit for short-term-inaccuracy.

These descriptions are based on the experience that long-distance connections in communication networks generally pass a

number of network nodes (which can be taken as random samples of a great number) for a comparatively short connection time, t_m . Every trial to establish a connection is taken as an elementary event, and the phase-deviation, measured over one or more nodes at time instant t , as a time-dependent random variable or stochastic process with the parameters e.g. 1. and 2. In the above mentioned consideration, the extensive mathematical description of time-dependent parameter-variations has been omitted, because of $t_m \ll K^{-1}$, see e.g. reference¹⁹.

1. Long-term inaccuracy $t > K^{-1}$:

In the general case of a linear network of N oscillators, which is mutually synchronized according to the phase-averaging principle, phase locked loops with an integral characteristic are excluded for the sake of drift stability. The averaged long-term phase-error between two network nodes (i and j) can be computed from the systems of equations (2) which leads to

$$\left\langle \frac{|p_i(t) - p_j(t)|}{C(N)} \right\rangle = \left\langle \left| \frac{\Delta f_0}{K} - f \delta t(t) \right| \right\rangle = \frac{\langle \Delta t \rangle}{T} \quad (8)$$

in which the pointed brackets (...) are used as an indication of ensemble-averaging. $C(N)$ represents a function of the weighting factors a_{ij} and the network configuration; Δf_0 represents the difference of the oscillator frequencies f_{j0} and f_{i0} in the off-control state; $f = T^{-1}$ represents the network frequency; and $\delta t(t)$ represents the difference between the averages of the propagation-delay-time variations $\Delta \tau_{ij}$ and $\Delta \tau_{ji}$, which are accepted by the nodes i and j , respectively. Concerning slight differences of synchronization methods see e.g. reference^{9,23}. As an estimated value, $\langle \Delta t \rangle$ represents the long-term phase-error according to Δf and δt as estimated means $\langle \Delta f \rangle$ and $\langle \Delta t \rangle$, respectively, or worstcase values. Detailed estimations have shown that the long-term time interval error is determined only by $\Delta f/K$. Therefore, the normalized fractional time deviation becomes:

$$\frac{\langle \Delta t \rangle}{T} = \frac{T}{T} \frac{\langle |\Delta f_0(t)| \rangle}{K} \approx \frac{\langle \Delta f(t) \rangle}{T} \quad (9)$$

Figure 8 confirms the fact that the effects of propagation delay-time variations Δt on the frequency error Δf of mutually-synchronized networks could be made as small as desired if K_0^{-1} is increased without changing to a DES. But, in the same way, the fractional time error $\Delta t/T$ increases if K_0^{-1} decreases.

2. Short-term inaccuracy

After discussing the stability of a generalized system without external excitation, we shall now evaluate the systems of equations (3) in the frequency domain $s = j\omega_b = j2\pi f_b$, that is, for undamped excitation and with elimination of terms concerning frequency drift. This yields, in correspondence to (4), the expression

$$\frac{P_i(j\omega_b) - P_j(j\omega_b)}{C(N)} = g(j\omega_b) \left[\frac{\Delta f(j\omega_b)}{K \cdot H(j\omega_b)} - \delta P(j\omega_b) \right] \quad (10)$$

in which $g(j\omega_b) = g_j(j\omega_b)$ from equation (7). In the following, for reasons of simplicity, $H(f_b) = 1/(1+j2\pi f_b T_h)$ is assumed; and the transient is nonperiodic i.e. $KT_h = 0,25 < 0,5$, as shown in figure 5.

Equation (10) shall now be expressed as a power spectrum. One obtains, in the frequency domain

$$\Phi_{\Delta P}(f_b) = C^2(N) |g(f_b)|^2 \left[\frac{\Phi_{\Delta F}(f_b)}{|K \cdot H(f_b)|^2} - \Phi_{\delta P}(f_b) \right] \quad (11)$$

in which $\Phi_{\Delta F}(f_b)$ is the Power Spectral Density (PSD) of the non-drifting freerunning frequencies, and $\Phi_{\delta P}(f_b)$ that of the non-drifting and weighted phase variations in Hertz²/Hz and Timeslot²/Hz, respectively.

To yield a first approximation, we shall, considering on the basis of work which has already been done^{13,14,15,16,17,18}, assume for $\Phi_{\Delta F}$ as well as for $\Phi_{\delta P}$ a white spectrum with a PSD of σ_F^2/f_{NF} or σ_P^2/f_{NP} , respectively, at least over the range of the noise bandwidth $f_N = K/4$ of $g(f)$. In this case, σ^2 is the finite power of variation of the process in the time domain. Thus, equation (11) takes, after integration to infinity in either direction, the form

$$\left[\frac{\sigma_{\Delta P}}{C(N)} \right]^2 = \left[(f'_{N0}/f_{NF}) \frac{\sigma_F^2}{K^2} + (f_{N0}/f_{NP}) \sigma_P^2 \right] = \frac{\langle \delta t^2 \rangle}{T^2} \quad (12)$$

which corresponds to the square of the ensemble average of the time deviations, $\langle \delta t^2 \rangle = \langle (\Delta t - \langle \Delta t \rangle)^2 \rangle$. In this expression, f'_{N0} represents the noise bandwidth of $g(f)/H(f)$. With $f_{N0} = K_0 K_0 / 4 = f'_{N0}$ we yield the expression

$$\frac{\langle \delta t^2 \rangle}{T} = \left[K_0^{-1} \left(\frac{K_0}{K_{0F}} \right) \cdot \left(\frac{\sigma_F}{K_0} \right)^2 + K_0 \left(\frac{K_0}{K_P} \right) \sigma_P^2 \right]^{1/2} \quad (13)$$

which, in real communication networks with a finite number of outputs (N), can be approximated only by an estimation value. Equation (13) expresses an opposition of that part of variations caused by frequency- or phase-noise, in dependence upon the slope characteristic of the phase discriminator, K_0 . This fact has been plotted in figure 9 for various products $K_F = (K_0/K_F)^{1/2} \cdot \sigma_F/K_0$ or $K_P = (K_0/K_P)^{1/2} \sigma_P$. For the final evaluations, it is assumed, that in equation (13) one or the other portion tends to dominate and, therefore, from

$$\frac{\langle \delta t^2 \rangle^{1/2}}{T} = \frac{1}{T} \frac{\langle \delta t^2 \rangle^{1/2}}{T} = \frac{1}{T} (K_0^{-1} K_F^2 + K_0 K_P^2)^{1/2} \quad (14)$$

either $K_0 K_F^2$ or $K_0 K_P^2$ as additional phase-variations can be computed independently.

Figure 10, basing on datas listed in table 2, gives a final summary of the long- and short-term behavior of mutually-synchronized (DES; SES) or directedly-synchronized (MSS) communication networks:

1. Long-term behavior

The fractional frequency inaccuracy of asynchronous quartz oscillators, with an assumed linear drift of frequency with respect to time, increases up to infinity.

Commercial Cs-atomic frequency standards show, in the asynchronous operation mode, a constant frequency inaccuracy. In the synchronous operation mode, the fractional time inaccuracy converges towards zero if the MSS synchronization method with a vanishing phase error is applied. On the other hand, the time inaccuracy of mutually-synchronized quartz oscillators remains constant with a distinct increase in quality compared to the asynchronous mode and a slight superiority of DES over SES.

2. Short-term behavior

The frequency inaccuracy of asynchronous oscillators is generally less than that of synchronous networks, because the phase variations of the transmission systems deteriorate the short-term stability of the oscillators. The fractional time error converges with increasing t towards the integral of spectrally-weighted inconstancy of the oscillators and phase variations of the transmission paths, according to equations (11) and (12). The DES shows a greater time error, for the feedback path yields an additional variation contribution. For reasons of comparison, the time interval error TIE, recommended by the CCITT for international communication systems²⁰, has been outlined in figure 10. Corresponding recommendations for FDM-networks are considered desirable^{21,22}. Nevertheless, short- and long-term behavior of synchronized TDM networks can be described by the indicated fractional time errors and, according to figures 8, 9 and 10, can be interchanged within certain limits. The coupling of synchronous networks yields a system-dependent, limited accumulation of time- and frequency errors, a fact that has not been considered in figure 10 but determines the structure constant $C(N)$ of equation (8) and (10). Field trials concerning both Directed- and Mutual synchronization, which are planned for the telephone network of the Federal Republic of Germany²⁴, are expected to contribute further guiding datas for more precise considerations.

Acknowledgement

The author wishes to express his thanks to Mr. Dr.-Ing. E. Steiner of Siemens AG for helpful discussions and suggestions, and to both the Deutsche Forschungsgemeinschaft and Siemens AG München, who have sponsored these studies. The author is sincerely indebted to his assistants, Dipl.-Ing. B. Müller and Dipl.-Ing. Chr. Jackisch, for their help in making evaluations and printing this paper.

References

- 1/ G.P. Darwin, R.C. Prim
Synchronization in a System of
Interconnected Units
U.S. Patent No. 2.986.723, May 30, 1961.
- 2/ J.P. Runyon
Reciprocal Timing of Time-Division
Switching Centers
U.S. Patent No. 3.050.586, Aug. 21, 1962.

- /3/ H.L. Hartmann
Synchronisierung integrierter PCM-Netze durch digital gesteuerte Phasenmittelung NTZ 7 (1968), H 9, S. 533 - 539.
- /4/ M.B. Brilliant
The Determination of Frequency in Systems of Mutually Synchronized Oscillators
Bell Syst. Techn. Journ. 45 (1966), Dec, pp. 1737 - 1748.
- /5/ J. Yamato, M. Ono, S. Usuda
Synchronization of a PCM Integrated Telephone Network
IEEE Trans CT-15 (1968) Febr, pp. 1 - 11.
- /6/ W. Herold
Synchronisation digitaler Fernmelde-netze durch Phasenmittelung mit Stell-größenübertragung
Dissertationsschrift, TU-München, 2. 3. 1972.
- /7/ A. Gersho, B.J. Karafin
Mutual Synchronization of Geographi-cally separated Oscillators
Bell Syst. Techn. Journ. 45 (1966), Dec, pp. 1689 - 1704.
- /8/ H. Inose, H. Fujisaki, T. Saito
Theory of Mutually Synchronized Systems
Abstracts Journal IECEJ 49 (1966) No. 4, pp. 19 - 21.
- /9/ O. Karl
Synchronization Procedures for Inte-grated PCM Networks using the Phase-Averaging Principle
NTZ 23 (1970), H 8, pp. 402 - 411.
- /10/ M.R. Miller
A Survey of Theoretical Studies of Digital Network Synchronization Systems
Proceedings of 1972 International Zürich Seminar, pp. D1(1) - D1(7).
- /11/ E.P. Graf, B. Walther
Generation of Base Band Frequencies for FDM and TDM Telecommunications
Proc. Frequ. Contr. Symp., this issue.
- /12/ CERN
Mésure des variations de distance de Symphonie par réception des signaux de Television
Note No. 75/254 DSES/SP/ST/VF
Toulouse, 23. 12. 1975.
- /13/ C.J. Byrne, B.J. Karafin, D.B. Robinson
Systematic Jitter in a Chain of Digital Regenerators
Bell Syst. Techn. Journ. 42 (1963) Nov, pp. 2679 - 2714.
- /14/ D.L. Duttweiler
Waiting Time Jitter
Bell Syst. Techn. Journ. 51 (1972) Jan, pp. 165 - 207.
- /15/ H.L. Hartmann, W.G. Janko, G. Heuse.
Akkumulation des Phasengeräusches in Stuffing-Multiplexsystemen
NTZ 28 (1975), H 2, S. 47 - 53.
- /16/ B. Müller
Messung der Langzeit Phasenschwankungen in Stuffing-Multiplexsystemen für PCM- und Datensignale
NTZ 29 (1976), H 4, S. 307 - 313.
- /17/ CCITT
Line Jitter in 2048 And 8448 KBit/s Com Sp. D-No. 288-E, October 1975, pp. 1 - 12.
- /18/ D.L. Duttweiler
The Jitter Performance of Phase-Locked-Loops Extracting Timing From Baseband Data Waveforms
Bell Syst. Techn. Journ. 55 (1976) Jan, pp. 37 - 58.
- /19/ L.S. Cutler, C.L. Searle
Some Aspects of the Theory and Measure-ment of Frequency Fluctuations in Frequency Standards
Proceedings of the IEEE, Vol. 54 No. 2 (1966) pp. 136 - 154.
- /20/ CCITT
Plesiochronous Operation of Internation-al Digital Links
Recommendation G. 811, AP VI-No. 66-E, pp. 190 - 194.
- /21/ P. Kartaschoff
Frequency Control and Timing Require-ments for Communications Systems
31st Annual Symposium On Frequency Control; this issue.
- /22/ E. Steiner
Carrier Supply Using Phase-Locked Loops for Higher-Order Frequency Division Multiplex Systems
NTZ (1972), H 11, pp. 523 - 527.
- /23/ H.L. Hartmann
The Single-Ended Principle of Digitally Controlled Phase Averaging for the Synchronization of Communication Net-works
NTZ 28 (1975), H 12, pp. 421 - 426.
- /24/ W.R. Slabon
Results of Investigations for the Clock Frequency Control and Distribution Systems in Digital Telephony and Data Networks of the Deutsche Bundespost and Future Plans
31st Annual Symposium On Frequency Control; this issue.

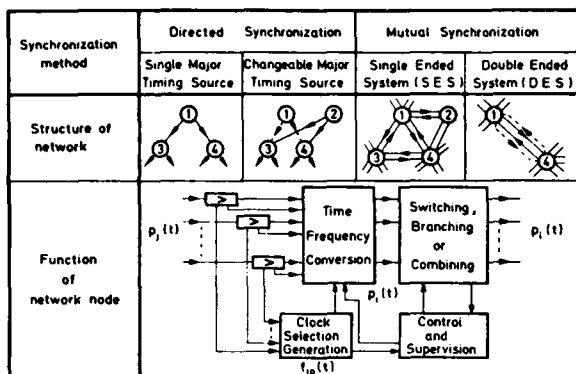


Table 1: Basic synchronization methods for communication networks

Method	K_D^{-1}	$\sqrt{K_D/K_P}$	σ_p	$\langle \delta t^2 \rangle^{1/2}$ $t = 1 \text{ sec}$	$\Delta t / K_D \text{ day}$	$\langle \Delta t \rangle$ $t = 1 \text{ day}$
SES	256	$3,2 \cdot 10^{-2}$	0,1	$9,8 \cdot 10^{-11}$	$5 \cdot 10^{-4}$	$7,2 \cdot 10^{-13}$
DES	8	$7,1 \cdot 10^{-2}$	$0,1 \sqrt{2}$	$1,7 \cdot 10^{-9}$	$\frac{1}{2} \cdot 2,5 \cdot 10^{-4}$	$5,7 \cdot 10^{-15}$

Table 2: Datas for first approximations

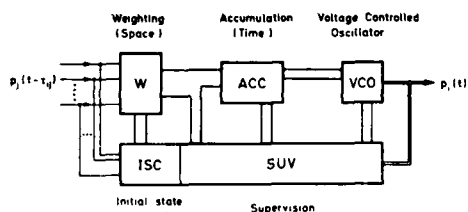


Figure 3: Generalized control loop for network synchronization

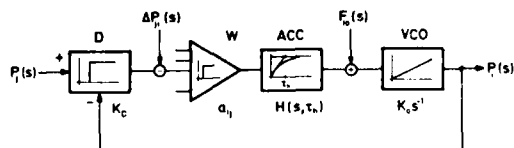


Figure 4: Spectral equivalent of generalized control loop

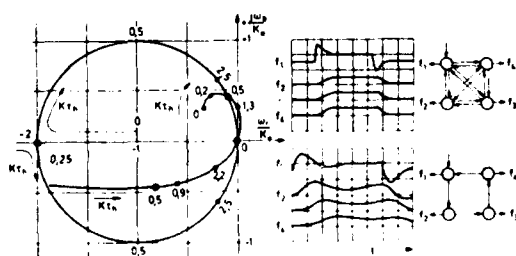


Figure 5: Polar plot and time behavior of the system for data of Table 1, parameter Kt_n

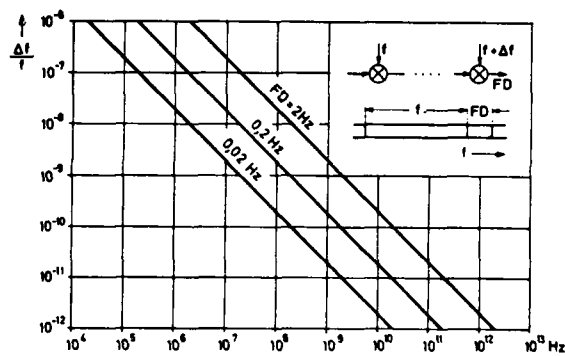


Figure 1: Frequency inaccuracy versus carrier frequency and frequency-deviation (FD)

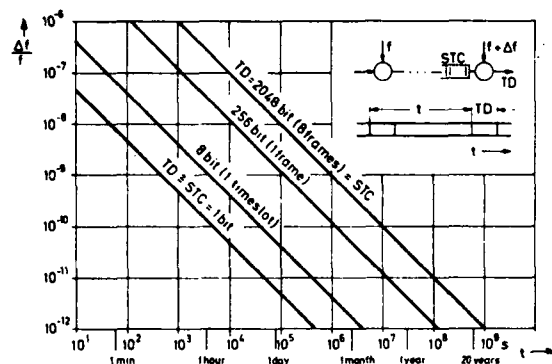


Figure 2: Frequency inaccuracy versus time between slips and time deviation (TD = Storage Capacity)

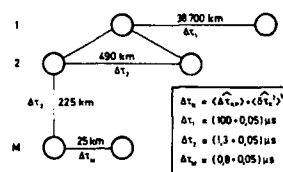


Figure 6: Reference communication network

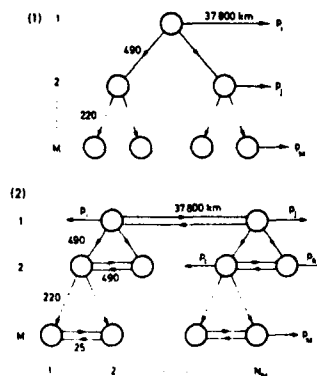


Figure 7: Master, slave, and mutually synchronized master-slave basic synchronization schemes

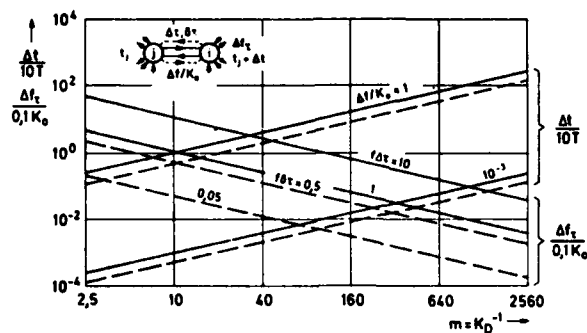


Figure 8: Long-term frequency modulation Δf_L and time-deviation Δt versus discriminator gain K_D ; —: SES; - - -: DES

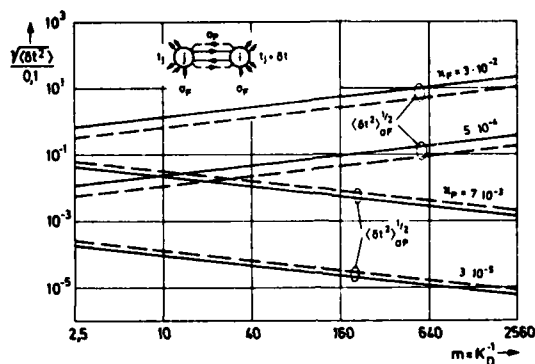


Figure 9: Short-term phase variation $\langle \delta \phi \rangle^{1/2}$ due to line jitter σ_p and frequency-noise σ_f versus discriminator gain K_D

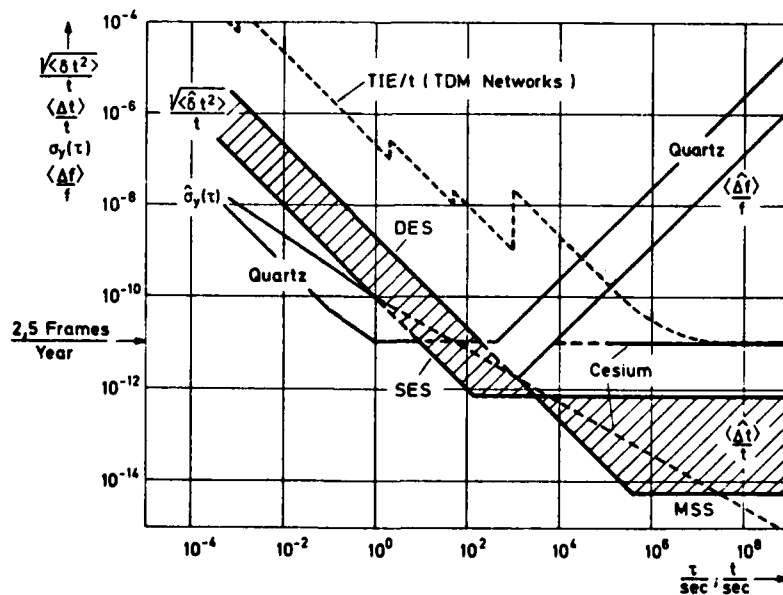


Figure 10: Approximated long- and short-term behavior of different frequency sources and synchronized networks

FREQUENCY CONTROL AND DIGITAL NETWORK SYNCHRONIZATION*

Marc I. Spellman
J. Bibb Cain
David B. Bradley

Harris Electronic Systems Division
Melbourne, Florida

Summary

Network timing is a fundamental requirement existing in communications networks which multiplex and switch digital signals. The network timing requirement is essentially fulfilled if the instantaneous nodal clock frequency at each node in the network is synchronous to all other nodal clocks in the network. This paper deals with four timing techniques, with their similarities and differences, and with their relative strengths and weaknesses in fulfilling the timing requirement in a switched, digital military communications network such as the future Defense Communications System. A conclusion is presented as to the most appropriate timing techniques and is supported by both analysis and simulation results.

The four timing techniques evaluated were independent clocks, master-slave, mutual synchronization, and time-reference distribution. The first technique, independent clocks, is an undisciplined technique because each clock in the network runs independent of any control from a reference. The remaining three techniques – master-slave, mutual synchronization, and time reference distribution – are disciplined techniques. In all of the disciplined techniques, each of the nodal clocks (with the exception of the master clocks in the time reference distribution and master-slave techniques) is locked to a reference signal. One of the principal differences between the techniques is the choice of the reference signal. This difference has direct impact on the capability of the technique to accommodate link-delay variations without deleterious effects on nodal clock frequencies. A second difference between the techniques is how each adapts to a loss of the reference signal at a disciplined node. This difference is manifested in the ability of each technique to keep the frequency of the network clocks closely aligned after such a failure. Other differences in techniques are the capabilities of providing precise time and clock error correction, and provides precise time, is the technique that should be employed at the major nodes of the network.

The analysis and simulation results that will be presented indicate that the time reference distribution technique, which distributes timing in a tree structure, includes double-ended control and adaptive reorganization, has an inherent separation of clock error measurement and clock error correction, and provides precise time, making it the technique that should be employed at the major nodes of the network.

Introduction

In order to ensure acceptable performance in a digital communications network, the timing or synchronization subsystem of that network must at the very least ensure that each clock in the network runs at the same average frequency. This, however, is only a basic requirement for the synchronization subsystem. In order to provide superior performance, additional properties should be provided. Each clock should run at, or extremely near, the same instantaneous frequency. The effect of path-delay variations on nodal clocks should be minimized. Network timing should be maintained in the face of stress situations. These additional properties are very significant in the type of network that is considered in this paper; that is, a worldwide, switched, digital military network such as the future Defense Communications System (DCS).

The key features associated with the military nature of the network are first, the predominance of secure or encrypted traffic traversing the network along with the resulting importance of maintaining bit-count integrity and second, the network stress not only from normal equipment failures common to all systems but from overt action taken by an enemy. The fact that it is a worldwide network ensures the linking together of numerous links as well as the inclusion of several different types of links (satellite, cable, LOS microwave, tropospheric scatter, etc.).

In this paper four techniques, master-slave, independent clocks, mutual synchronization and time reference distribution (each of which is defined in detail in the following paragraph) are compared in order to determine the most appropriate technique to use in the system outlined above. The conclusion drawn from a preliminary comparison of the techniques indicates that each of the four techniques could satisfy the basic synchronization requirement of ensuring that each clock in the network runs at the same average frequency. In truth, the viability of these techniques is confirmed by the fact that each of these techniques has either been implemented or is planned to be implemented in at least one digital communications network.

As a result of the equally high degree of potential capability associated with each of the techniques, a specific solution methodology was defined in order to draw quantitative comparisons between the alternate techniques. In general, the bit-slip rate for a given buffer size was selected as the key quantitative measurement parameter; that is, for equivalent buffer size, which technique would provide the smallest number of bit slips over a given time interval. Buffers are required even in a theoretically ideal timing subsystem where every clock in the network runs at precisely the same instantaneous frequency. These buffers accommodate the expected path-delay variations on a single link. As the timing subsystem performs "less optimally," the size of each buffer in the network must be enlarged, if no increase in bit slips is to occur. This is true since the buffers must now accommodate differences in instantaneous nodal clock frequencies as well as single link phase perturbations. Although buffering is relatively inexpensive, it is certainly desirable, "ceteris paribus," to minimize buffer size, because, at the very least, increases in buffer sizes result in increases in transmission delays.

The solution methodology followed here defined both a number of desirable characteristics that enhance overall subsystem utility and a timing-subsystem test bed consisting of a set of possible network configurations and stress scenarios. Next, the degree to which each technique provided the desirable characteristics in the test-bed situation was evaluated.

Overview of Candidate Techniques

The four candidate techniques evaluated consisted of independent clocks which can be characterized as an undisciplined- or an unlocked-clock approach and master-slave, mutual synchronization, and time reference distribution all of which are locked or disciplined clock techniques.

Independent Clocks

The independent clock approach is a synchronization approach using unlocked clocks which requires a very stable independent clock at each node. Data traffic arriving from other nodes in the network having independent clocks must be buffered. This buffer must be sized to accommodate the possible frequency offsets between the independent clocks and variations in frequency due to transmission phenomena.

If the nodal clocks are set to run at a nominal data rate of f_0 , and the data rate of any two nodal clocks is within Δf_0 of f_0 , then the maximum difference in data rate between any two nodal clocks is $2\Delta f_0$. Assuming that a buffer size of m bits is chosen (set initially at half full), the time required for a buffer overflow is at least

$$T = \frac{m^2}{2\Delta f_0} \quad (1)$$

letting

$$S = \frac{\Delta f_0}{f_0} \quad (2)$$

denote the long-term clock stability factor, (1) becomes

$$T = \frac{m}{4Sf_0} \quad (3)$$

Then, using Equation (3), the buffer size which guarantees that the time between overflows or underflows is at least T seconds can be computed. If one has an atomic frequency standard with $S = 10^{-11}$, then using Equation (3), one finds that a buffer size of 1040 bits is required (at $f_0 = 10$ Mb/s) to guarantee a period of 30 days between buffer overflows. Thus, one can see that buffer sizes may be significant with this approach, even using atomic clocks.

There are two disadvantages that the independent clock technique has vis-a-vis any of the disciplined clock approaches. First, because clocks are not disciplined, ultimate inherent accuracy must be demanded from individual clocks. Thus, the number of cesium clocks required in the network will be significantly larger than any of the locked-clock approaches. This results in a significant cost impact. Second, despite the highly accurate clocks, periodic bit slips must be accepted since clocks are not disciplined. The complete independence of the nodal clocks does, however, insulate the network performance from local failures and/or vulnerability.

Master-Slave

Master-slave is a very straightforward disciplined approach. Timing is distributed through a master-slave tree. The nodal clock is derived by locking a very narrowband phase-locked loop to the received bit timing from an incoming trunk. This approach has been used in many of the commercial systems that have already been built (Western Union, Datan, and AT&T). Generally, a hierarchical approach with a number of diverse routes for timing distribution has been used. This gives an added measure of survivability, since it minimizes the probability that a given node will be forced to operate asynchronously due to a failure in the particular link from which it derives timing. An added degree of sophistication that is feasible with respect to alternate routing is adaptive reorganization. This feature is compatible with either master-slave or time-reference distribution since both are based on hierarchical tree structures. However, because it requires a duplex control link and because such a link is already part of the time-reference distribution system and not a part of the master-slave system, this feature has been incorporated in the time-reference distribution approach.

Mutual Synchronization

Mutual synchronization is a synchronous technique where the reference for the local nodal clock is an average of the timing signals derived from each of the incoming trunks. Obviously, the loss of a single link will not necessitate a network reorganization as it might in a master-slave system. However, unlike a master-slave system, a link perturbation on any of the incoming trunks can have an effect on the nodal clock. In general, it can be said that a common average frequency for all nodes in the network is achieved in the mutual synchronization system at the expense of increased variations in instantaneous frequencies between nodes.

Time Reference Distribution

Time reference distribution (TRD) is more closely related to a master-slave system than any other technique. In fact, it can be described as a double-ended master-slave system that is referenced to some standard of precise time and includes an adaptive reorganization capability and a separation of error measurement and error correction. This point is worthy of elaboration. In order to elaborate on this statement, four points must be discussed. These are:

- Double-ended versus single-ended
- Adaptive reorganization
- Reference to precise time
- Separation of error measurement and error correction

As the techniques have been defined, master-slave is a single-ended system and TRD is a double-ended system. The single-ended approach uses a reference timing signal local to the clock being disciplined. That is, the reference signal is the timing signal derived from one of the received trunks. This signal is essentially the timing signal generated at some distant node, corrupted by the link variations inherent in the media and corrected by the appropriate filtering in the receive node. Such a system does not require an overhead link for timing information.

The double-ended system attempts to lock the nodal clock directly to the timing signal generated at the distant disciplining node. This is accomplished in TRD by measuring the phase offset between clocks via a time transfer technique, as illustrated in Figure 1, and by correcting the disciplined clock so that the measured phase differences between clocks is driven to zero. The time transfer technique requires an overhead link from the disciplining node to the disciplined node.

The adaptive reorganization capability included in the time reference distribution approach addresses the problem of how nodes within the network attain a timing reference after the primary reference has gone down. The adaptive reorganization approach essentially results in a real-time rating of all the potential timing reference sources available for a node. This rating is based on link qualities, clock types, and timing distribution paths to the ultimate master. The technique guarantees that the best available reference for each node is selected in case of failure. The requirement placed on the network for this service is an overhead channel required for the distribution of link and nodal rankings. The alternative to the adaptive reorganization scheme is a fixed reorganization scheme that can result in deleterious timing results in the face of certain combinations of failures.

A valuable by-product of the TRD approach results if the network master is referenced to a source of precise time such as UTC at the Naval Observatory. If this is done, the network will now have available at each node a source of precise time. A double-ended system (such as TRD) is the only type that can provide this characteristic. The resulting benefits are discussed later.

The overhead link required in the TRD approach for time transfer and adaptive reorganization can also be used for the transmission

of measured but uncorrected clock errors from node to node. The TRD approach provides a natural division between error measurement and error correction. If the time interval for error correction is long, the measured, but uncorrected, error on a series of tandem links may be provided via overhead to nodes lower in the hierarchy. This permits the simultaneous correction of all clocks in the network directly to the master clock; thus, all clocks are brought more quickly to synchronism.

Criteria for Evaluation

In order to compare the candidate techniques in a definitive and consistent manner, a test bed was defined which consisted of the following three areas:

1. A set of desirable characteristics which, if possessed by a candidate technique, would enhance the overall utility of the synchronization subsystem.
2. A set of network configurations each of which was representative of the type of network the timing subsystem might serve.
3. A set of stresses through which the candidate techniques should hopefully continue to function.

The goal of the evaluation of the techniques was to determine which of the candidate techniques would most consistently provide the desirable characteristics while operating through stress situations in the representative network configurations.

Desirable Characteristics

A number of characteristics have been identified that, if possessed by a timing subsystem, would be considered advantageous from a system viewpoint. These desirable characteristics, although not necessary for the provision of a minimal level of performance, do enhance the overall utility of the timing subsystem. The desirable characteristics are survivability, limited error propagation, compatibility with other global timing subsystems, precise time availability, stability, monitorability, minimum susceptibility, flexibility, interoperability, and maximum availability. Below, each of these characteristics is defined and is justified as being important in the overall subsystem evaluation.

1. Survivability. Survivability indicates the degree to which the timing subsystem continues to perform its function (to provide network timing) during periods of stress. Obviously, a military communications network must provide dependable communications for its users during stress situations. The frequency offset between nodes during stress is taken as the measure of survivability of the timing subsystem. If this offset increases in a stressed environment, then bit slips increase which lead to misaligned frames in multiplexers and switches and a need to resynchronize cryptos. If the slip rate becomes excessive, data traffic will degrade significantly.

2. Error Propagation. It is desirable to prevent disturbances (such as stresses or path-delay variations) in one part of the network from propagating and influencing the phase of nodal clocks in other parts of the network.

This is an important characteristic because in preventing the propagation of perturbations in nodal frequencies from node to node throughout the network, one maximizes the stability of each nodal clock and minimizes the buffer size required to achieve a desired slip rate. The parameter that will be used as a measure of error propagation is the offset in phase of each nodal clock from their steady state values as a result of certain network stresses.

3. Compatibility With Other Global Timing Subsystems. This characteristic refers to the degree with which the network frequency and/or precise time agrees with that provided by other global Navigation/Timing Systems such as Loran-C, Omega, and GPS. Such agreement enhances the availability of the timing function since it provides functional redundancy.

4. Precise Time Availability. Precise time can be defined as the correct absolute time as kept by a standards laboratory such as the USNO. The timing subsystem could be designed to acquire and disseminate precise time from the USNO within some margin of error. This capability is desirable for the following reasons:

- a. The network is automatically referenced to UTC and has the advantages of better interoperability with other networks referenced to UTC and compatibility with other global timing systems.
- b. Precise time is useful in assisting synchronization of spread spectrum equipment and in providing synchronization for TDMA networks.
- c. Precise time would be more widely available to those users of precise time within the government. This will simplify their problems of updating their precise time standards.

5. Stability. The timing subsystem must be a stable system (in a feedback control sense).

This is critical because if the system were not stable, nodal clocks could be continually pulled off in one direction or simply be unable to maintain network synchronization. The results of such an occurrence would be buffer overflows and therefore, continual bit slips.

6. Monitorability. Monitorability is the level to which the timing subsystem can provide data used in assessing its own performance. It is apparent that monitorability is valuable at least to the extent that an option such as a redundant piece of equipment or an alternate link is available and can be utilized as a result of decisions made on the basis of monitored information.

Monitorability is a significant characteristic because it can lead to greater availability. If potential failures can be identified prior to failures actually occurring, these situations can be remedied on an off-line basis. The result is therefore both an increase in MTBF and a reduction in MTTR and therefore, a greater availability.

7. Susceptibility. Susceptibility is the degree to which the timing subsystem may be spoofed, jammed, or disrupted by an enemy via electronic means, as well as disrupted due to natural phenomena.

Because of the critical nature of the military communications mission, it is apparent that it could be a focal point of enemy action during conflict. Therefore, susceptibility is a key characteristic. Any timing subsystem that by its nature increases the susceptibility of the network as a whole represents a significant handicap. Slip rate will again be the measurable quantity in evaluating susceptibility. The evaluation consists of measuring the rate of accumulation of phase error between the nodal clocks (slip rate is directly proportional to the rate of accumulation of phase error between the nodal clocks at two nodes and inversely proportional to the buffer size) under a variety of possible stress scenarios.

8. Flexibility. Flexibility is the degree to which the timing subsystem is compatible with the orderly implementation, growth and extension of the network. Such a feature is especially desirable in a military communications network which must often provide at a later date, service and capability unknown at the present.

9. Interoperability. Interoperability is the degree to which the timing subsystem influences the traffic interface between the network it serves and other military and commercial communications systems. Such interoperability is significant because of the inevitable use of facilities external to the network itself.

10. Availability. Availability which is the quotient of mean-time-before-failure and the sum of mean-time-before-failure (MTBF) and the mean-time-to-repair (MTTR) is a significant characteristic for all major systems, and its importance is apparent for the timing subsystem of a military communications network.

Network Configurations

A number of network configurations were used in the computer simulation that was required to evaluate desirable characteristics such as error propagation and survivability. Two reasons precipitated the use of several different configurations rather than a single network configuration. First, the nodal phase and frequency offsets in any of the disciplined techniques are dependent on the network configuration. Second, the desire to keep the results of the simulation as general as possible. Two of the connectivity patterns used are illustrated in Figures 2 and 3. Figure 2 was called the Dumbbell Node Configuration because of its peculiar shape. It consists of two relatively congested communications areas interconnected by a relatively small number of links. The Meandering Node Configuration (Figure 3) is so named because of the backbone node path consisting of Nodes 5, 8, 14, 18, 20, 23, 24, and 29 which are interconnected to a meandering path similar to a river.

Stress Scenarios

Three types of network perturbations were included in the evaluation. First, normal phase-delay variations which can be expected on various transmission media; second, changes in clock frequencies which can result from various clock failures; third, loss of a node and/or link with the resultant transience in network frequency.

Computer Simulation

Goals

The computer simulation was generated in order to provide a means for studying the reaction of a complex interconnected network to various disturbances and stresses. It is mainly due to the numbers of nodes and links that must be used in a representative model that a computer simulation is deemed necessary. The two- or three-node case can be analyzed without such aids. Normal disturbances include clock jitter, clock drift, initial frequency and phase offsets, and normal link-delay variations. Abnormal stresses include clock failures, node failures, link failures, and abnormal link-delay variations. The simulator can thus be used for comparing the performances of various timing techniques. The major parameters used in such comparisons include the following:

1. The speed with which the network disposes of a transient condition.
2. The degree of closeness that the nodal clock frequencies are held together.
3. The amount of phase error that can be found at the various nodes.

During the period of recovery from a disturbance occurring somewhere in the network, the timing subsystem is more vulnerable to any other disturbances that might occur. During transient conditions, error detection equipment has a degraded frame of reference and reconfiguration apparatus is more susceptible to error in the worst-case and, at best, is subject to a delayed response. Thus, shorter transient time is a desirable feature. As stated earlier, the ideal condition is that all clocks in the network remain exactly together. By providing receiving buffers on each link, some tolerance to temporary differences in clock frequencies may be obtained. However, any two communicating nodes must have an average difference frequency of zero if finite size buffers are used and loss of data is avoided. Thus, clocks held closely together are desirable attributes. Small phase errors at any instant of time are desirable because phase errors represent a difference in clock time readings and are stored potential which may be converted to a kinetic change in clock frequencies as a result of some future disturbance.

Timing Technique Models

The simulator consists of models of the timing techniques, transmission link media models, and statistical clock models. For the timing technique models, only the three disciplined techniques were

considered since the performance evaluation of the independent clock technique did not require a simulation model. Each of the disciplined clock technique models consisted of the following segments: a description of how the reference information was to be disseminated; an organization, reorganization procedure for the network; and an analytical description of the processing of the timing information at the node. In the case of the master-slave technique, reference information which is obtained from the bit transitions of the incoming data is distributed via a subset of links which is superimposed on the communications network. Reorganization of the network is based on a predetermined ranking of all available references at each node. Processing of the timing information is accomplished via the phase-lock loop which is shown in Figure 4. The closed loop has a transfer function which may be written as

$$\frac{\Phi_O(s)}{\Phi_I(s)} = \frac{K_V F(s)}{S + K_V F(s)} \quad (4)$$

The loop type included in the model is a type-two loop which is obtained by using an integral plus a proportional loop filter; i.e., $F(s) = (s + a)/s$. The closed loop transfer function then becomes

$$\frac{\Phi_O(s)}{\Phi_I(s)} = \frac{2\xi\omega_N s + \omega_N^2}{s^2 + 2\xi\omega_N s + \omega_N^2} \quad (5)$$

where

$$\omega_N = \sqrt{a K_V} \quad \text{and} \quad \xi = \sqrt{K_V/4a}$$

Two principal differences of this type of loop as compared to a type-one loop are that the integrator in the loop filter provides "frequency memory" for use when "coasting" without a reference and that the loop bandwidth can be changed as an acquisition aid to speed up acquisition from large frequency offsets. The loop parameters used in tracking do not impose a limit on the acquisition time as they do when using a type-one loop. This will allow the use of very narrowband loops while still retaining good acquisition performance. This two bandwidth approach is included in the model. An extended-range phase detector improves the performance of the loop significantly in this application by allowing the use of narrower bandwidths, and it also has been included in the model.

The loop parameters selected for the acquisition and tracking modes are shown in Table 1.

Table 1

Tracking $\xi = 4$, $\omega_N = 5.6 \times 10^{-5}$ rad/sec

For a drift rate of $1 \times 10^{-10} \times f_0$ per day, these parameters give a steady-state phase error of $0.37 \mu\text{sec}$.

Fast Acquisition $\xi = 0.707$, $\omega_N = 0.007$ rad/sec

The loop will acquire from a combined fractional frequency offset of 2×10^{-8} and a phase offset of $2 \mu\text{sec}$ in 1800 sec.

For the mutual synchronization model, a type-one loop was selected. Stability analyses for single-ended mutual synchronization systems have shown that in order for a mutual synchronization system to be stable, it requires a closed-loop phase response that satisfies

$$\left| \frac{\Phi_O(j\omega)}{\Phi_I(j\omega)} \right| = |H(\omega)| < 1, \omega > 0 \quad (6)$$

It can be shown that type-two loops with filters of the form $F(s) = (s + a)/s$ never satisfy this condition (note that $|H(j\omega)| > 1$). This means that phase jitter at the natural frequency will be amplified each time it passes through the timing regeneration circuitry at a node. Since this is a closed-loop system of many interconnected PLL's, one can see that the amplification of phase jitter is the source of the instability. The selected loop parameters are shown in Table 2.

Table 2

$$\xi = 1, \omega_N = 1.67 \times 10^{-3} \text{ rad/s}$$

For a fractional frequency offset of 1×10^{-8} , these parameters give a steady-state phase error of $2 \mu\text{sec}$.

The TRD technique uses the same kind of structure for disseminating reference information as is used for the master-slave technique. However, it is a double-ended scheme; i.e., there is a two way exchange of information between adjacent nodes in a chain such that the effects of transmission delay variations are removed. In this manner, the nodal clocks may be held as closely together as two-way path-delay asymmetries and the measurement errors allow. The model assumes that a special overhead channel is utilized for carrying the reference information. A further difference between the TRD and master-slave techniques is that reference information concerning an estimated error between the ultimate reference source and each node in the chain is added in and passed on to each succeeding node in the chain. This information is not filtered before being utilized.

The TRD model also utilizes an adaptive, self-organizing scheme rather than a fixed, reorganizing scheme. The adaptive scheme included in the simulation model is based on a system of clock quality factors and link demerits.

The processing of timing information in the TRD technique that is included in the TRD model is analogous in some respects to the processing in the master-slave technique. The TRD error processor is implemented as a loop filter.

The error processing filter selected is shown in Figure 5. It is an integral plus a proportional filter. If timing errors measured in nS are fed to the filter, a correction term C in nS/s appears on the output. As indicated the proportional term performs the short-term phase correction while the integral term performs the correction for a frequency offset between the local clock and the master. The time constants for these two correction terms can be made considerably different by choosing a $\ll k$ (i.e., selecting a large value of ξ).

Two sets of filter parameters were determined, one set for cesium clocks and another set for quartz clocks. The selected parameters are given in Table 3.

Table 3

Cesium Clocks

Only tracking parameters are used

$$\omega_N = 1.12 \times 10^{-5} \text{ rad/s}, \xi = 2$$

$$\text{These give } K = 4.5 \times 10^{-5} \text{ S}^{-1} \text{ and } a = 2.8 \times 10^{-6} \text{ S}^{-1}$$

This gives a peak frequency error of 4.5×10^{-11} when a step of $1 \mu\text{sec}$ of phase offset is applied

Quartz Crystal Clocks

$$\text{Tracking } \omega_N = 1.12 \times 10^{-4} \text{ rad/sec}, \xi = 2$$

This results in $K = 4.5 \times 10^{-4} \text{ S}^{-1}$ and $a = 2.8 \times 10^{-5} \text{ S}^{-1}$. For a drift rate of 1×10^{-10} per day, these parameters provide a steady-state phase error of $0.09 \mu\text{sec}$.

$$\text{Fast Acquisition } \omega_N = 0.007 \text{ rad/sec } \xi = 0.707$$

$$\text{This results in } K = 0.001 \text{ S}^{-1} \text{ and } a = 0.005 \text{ S}^{-1}$$

Link Models

Since the nodes in the network can be connected by links using coaxial cable, line-of-sight microwave, satellite, or tropospheric scatter media, the phase variations for each had to be quantified. The models used were derived from the literature.^{2,3,4}

Statistical Clock Models

Three types of timing clocks were considered. They are the cesium, beam-controlled oscillator, the rubidium, gas-cell, resonator-controlled oscillator and the quartz crystal oscillator. No matter which of these types of clocks are used at any given time at any node, its essential properties may be described by a list of parameters which are controlled by random number generators. The parameter list contains four parameters for each nodal clock. These parameters include frequency, short-term stability, and accuracy or long-term stability and drift rate. Frequency refers to nominal frequency. This parameter is generated for each node from a random generator at the beginning of each simulation run and is modified during the run by the accuracy random-number generator or by the drift-rate parameter. At each simulation step, a short-term stability offset is generated and added to the current value of nominal frequency to provide the mean frequency for the simulation step. From vendor literature, it appears that short-term stability (up to approximately 0.4 second) is essentially the same for all three types of frequency standards if drift rates are not included for rubidium and quartz. This literature shows measurements for square root of the sum of the squares of successive differences in average frequency measurements for measurement intervals from 10^{-3} to 10^{+4} seconds. Thus, the short-term stability random number generator may sample from a normally distributed variable with a zero mean and a standard deviation equal to $\frac{\delta \Delta f}{f} (2, \tau)$, where τ corresponds to the simulation step interval.

These deviations are given as piecewise functions of τ for each clock type in Table 4 along with systematic drift for each clock.

Table 4

Quartz

$$-3 \leq \log_{10} \tau \leq -0.55 \quad \log_{10} \sigma = -\log_{10} \tau - 11.7049$$

$$-0.55 < \log_{10} \tau \leq 4 \quad \sigma = 7 \times 10^{-12} \text{ or } \log_{10} \sigma = -11.1549$$

$$\text{Systematic Drift } \pm 1 \times 10^{-10} / \text{day} \rightarrow 1.57 \times 10^{-15} / \text{sec}$$

Rubidium

$$-3 \leq \log_{10} \tau \leq -1 \quad \log_{10} \sigma = -\log_{10} \tau - 11.7049$$

$$-1 < \log_{10} \tau \leq 2 \quad \log_{10} \sigma = -0.5 \log_{10} \tau - 11.1549$$

$$2 < \log_{10} \tau < 4 \quad \sigma = 7 \times 10^{-13} \text{ or } \log_{10} \sigma = -12.1549$$

$$\text{Systematic Drift } \pm 1 \times 10^{-11} / \text{month} \rightarrow 3.858 \times 10^{-18} / \text{sec}$$

Cesium

$$-3 \leq \log_{10} \tau \leq -0.55 \quad \log_{10} \sigma = -\log_{10} \tau - 11.7049$$

$$-0.55 \leq \log_{10} \tau \leq 0.6 \quad \sigma = 7 \times 10^{-12} \text{ or } \log_{10} \sigma = -11.1549$$

$$0.6 \leq \log_{10} \tau < 4 \quad \sigma = -0.5 \log_{10} \tau - 11.04$$

$$\text{Systematic Drift } \text{None} \quad \text{Long-Term Stability } \pm 1 \times 10^{-11}$$

Typical Simulation Results

Some typical simulation results are presented in Figures 6, 7, and 8. These figures and their meaning are briefly explained, and then a summary of the simulation results is presented.

Figures 6, 7, and 8 each display nodal frequency-error plots at selected nodes in the dumbbell network configuration that result from initial transients plus normal link disturbances plus stress events. Figure 6 displays these results for the mutual synchronization technique; Figure 7 displays the results for the master-slave synchronization technique; and Figure 8 displays the TRD results.

Figure 6 for example shows the nodal frequency errors for Nodes 15, 17, and 25. The step change in the frequencies of $1 \times 10^{-8} \times f_0$ at time 10,000 seconds of Nodes 18, 21, and 25 caused Node 17's clock to change by $32 \times 10^{-10} \times f_0$ but had little effect on the nodes in the left half of the network. The change of $1 \times 10^{-8} \times f_0$ in clocks 13, 14, and 15 at 20,000 seconds had little effect on the nodes of the right half of the network. Node 16 acted as a buffer during the period from 10,000 seconds to 20,000 seconds. The net result of these two changes was almost zero by 40,000 seconds. At 40,000 seconds when Links 16-18 failed, Node 15's clock decreased in frequency by approximately $27.6 \times 10^{-10} \times f_0$. Loss of Links 17-18 at 60,000 seconds caused Node 17's clock frequency to decrease by about $17 \times 10^{-10} \times f_0$. The increase in delay on Links 14-16 at time 100,000 seconds caused almost no change in any of the three plotted nodes' frequencies. When Links 15-16 failed at time 140,000 seconds, Node 15 decreased in frequency by approximately $73 \times 10^{-10} \times f_0$. Failure of Node 17 at 180,000 seconds caused the clock frequency of Node 15 to decrease by approximately $17 \times 10^{-10} \times f_0$ and the clock frequency of Node 25 by approximately $11 \times 10^{-10} \times f_0$. The smaller change in Node 25's frequency is attributed to the larger number of nodes connected to 25. Failure of Node 24 at 220,000 seconds caused 25's frequency to momentarily decrease by approximately $32 \times 10^{-10} \times f_0$.

A tabular summary of the bulk of the simulation run results will be much more helpful and is shown in Table 5.

A link dropout or node failure results in changing references in the TRD and master-slave techniques. It results in reaveraging phase errors from the remaining references in the mutual sync technique. As a result of these types of stresses, the master-slave technique showed peak frequency error up to 3.25 parts in 10^{10} . The mutual sync technique showed peak frequency changes of from ten to twenty times that of master-slave while no variation in nodal frequency was experienced with the TRD technique due to these stresses.

The response of the TRD node to a VCO center frequency change was due to a time skew in the components that make up the total time reference signal with respect to the master at nodes further down the timing chain than where the stress occurred. This time skew in error components was a result of the method of implementation of the simulator but is not inherent to the TRD technique. This response does not die out in going further out the chain as with the master-slave technique because with TRD each succeeding node receives the same error due to the skewed samples.

The master-slave and TRD techniques experienced the same change in nodal frequency when switching from tracking to acquisition mode. However, this is strictly a function of the loop parameters and phase threshold at which the switch occurs.

Table 5

Disturbance	Peak Frequency Change ($\Delta f/f_0 \times 10^{10}$)			
	Nodes Away	Mutual Sync	Master-Slave	TRD
Link Dropout or Node Failure	1	Max - 65 Nom - 20-30	Max - 3.25 Nom - 2	0
	2	Max - 30 Nom - 3-10	Max - 2 Nom - 1	0
Change in VCO Center Frequency of 1×10^{-8}	1	10-33	81	7
	2	0-10	10	7
Switch From Tracking to Acquisition Mode	0	-	192 $[2(\delta_A \omega_{nA} - \delta_T \omega_{nT})\Delta\phi]$	192
Satellite Link Disturbance	1	Max - 7.5	0.6	0
	2	2.5-5.5		

Evaluation of Candidate Techniques

In order to compare the candidate techniques, their ability to provide each of the desirable characteristics delineated above was evaluated. The results are summarized below.

Survivability

Since timing subsystem performance is directly tied to slip rate, slip rate as a function of level of stress was chosen to be indicative of the survivability of each timing approach.

There are two parts to the problem of minimizing the slip rate in a digital network:

1. Keep the nodal clocks of all communicating node pairs synchronous as much as possible.
2. Minimize the frequency offset between the nodal clocks of asynchronous communicating nodes and minimize the amount of time they must communicate with asynchronous nodal clocks.

To put these comments in perspective, note that the independent clock technique uses asynchronous clocks at every node. The slip rate due to buffer overflows is strictly a function of the clock accuracy and the buffer size and will not change as a function of stress level. On the other hand, one could reduce the slip rate by slaving these clocks to a common master through some timing distribution hierarchy. This would make all nodal clocks synchronous and slips due to buffer overflows could be avoided entirely providing there were no link or node failures. If link or node failures occur, then some nodes may lose their reference to the master. Then some nodes with asynchronous nodal clocks may be communicating and could experience slips. Such nodes are called "asynchronous communicating nodes" and will remain asynchronous until they once again become locked to a common reference. It is very desirable that the timing approach minimize the period of time during which such nodes are asynchronous.

From these comments one can take issue with the common misconception that the independent clock approach offers the ultimate in survivability. One can argue that a disciplined technique with a fixed timing distribution network has better survivability because

nodes will not experience slips as long as they can obtain a reference, and when a reference is not available, they revert to a self-reference approach which has performance that can approach the independent clock technique. However, the improvement in using this approach is not great because of the significant probability that a node will lose its reference. This is the reason that hierarchical disciplined approaches have been criticized for having poor survivability. However, there is a better way for distributing timing. In fact, a dramatic improvement in survivability can be obtained by using an adaptive timing distribution network. This is due to the ability of each node connected to the network to find another reference after the loss of a reference. An analysis was carried out based on the expected time between bit slips in selected subnetwork configurations that supports this statement. The TRD approach which includes the adaptive timing distribution feature was therefore considered the most survivable approach.

Error Propagation

It is highly desirable to prevent perturbation in the frequency of one nodal clock from propagating and influencing the frequency of nodal clocks in other parts of the network. By minimizing such perturbations of nodal clocks, one maximizes the stability of each nodal clock and minimizes the probability of a bit slip.

The degree of error propagation for several types of disturbances was evaluated for each of the disciplined timing techniques. Four types of disturbances were evaluated. These were a link dropout, a step change in nominal VCO frequency, a step change in path delay, and sinusoidal path-delay variations. The disturbances were applied at one point in a network and the effect on other nodal frequencies was measured as a function of distance (in number of nodes) from the disturbance. The results are summarized in Tables 6, 7, 8, and 9.

Table 6. Error Propagation Due to a 0.65 μ s Step Change in Path Delay

Distance (Nodes)	Peak $\Delta f/f_0$		TRD
	Master-Slave	Mutual Sync	
1	3×10^{-10}	10^{-10}	—
2	10^{-10}	3×10^{-11}	—
3	6×10^{-11}	2.5×10^{-11}	—
4	4×10^{-11}	2.5×10^{-11}	—

Table 7. Error Propagation Due to a Sinusoidal Path Delay Variation on All Links With a Magnitude of 0.65 μ s Peak-to-Peak and a Period of 1 Day

Distance (Nodes)	Peak $\Delta f/f_0$		TRD
	Master-Slave	Mutual Sync	
1	2.5×10^{-11}	1.4×10^{-10}	—
2	5×10^{-11}	1.4×10^{-10}	—
3	7×10^{-11}	1.4×10^{-10}	—
4	9×10^{-11}	1.4×10^{-10}	—

Table 8. Error Propagation Due to a Link Outage

Distance (Nodes)	Peak $\Delta f/f_0$		TRD
	Master-Slave	Mutual-Sync	
1	1.7×10^{-10}	4.5×10^{-9}	6.5×10^{-11}
2	6.5×10^{-11}	8×10^{-10}	6.5×10^{-11}
3	4.5×10^{-11}	4×10^{-10}	6.5×10^{-11}
4	4×10^{-11}	4×10^{-10}	6.5×10^{-11}

Table 9. Error Propagation Due to a Fractional Frequency Step of 10^{-8} in Nominal VCO Center Frequency

Distance (Nodes)	Peak $\Delta f/f_0$		TRD
	Master-Slave	Mutual-Sync	
1	8×10^{-10}	1.4×10^{-9}	8×10^{-10}
2	6.5×10^{-11}	5×10^{-10}	8×10^{-10}
3	1.3×10^{-11}	5×10^{-10}	8×10^{-10}
4	1.3×10^{-11}	5×10^{-10}	8×10^{-10}

Compatibility With Other Global Timing Sources

Compatibility with other global timing sources was judged significant since this would result in greater redundancy for the timing function at each node.

Two points were considered when the alternative techniques were evaluated. First, did any of the techniques lend itself to an easier hardware implementation to achieve this compatibility? The conclusion was that all the techniques required essentially the same hardware. Second, would any of the techniques result in a smoother transition to the external timing source? (In this case, the term "smoother transition" refers to a minimum phase/frequency transient at the node transferring to the external source and in the network as a whole.) The conclusion here was that the TRD technique offers the potential for a smoother transition because of the likelihood that if the technique were implemented, both external and internal timing would be referenced to a common time standard such as UTC.

Precise Time Availability

If the master clock in the TRD system is referenced to a precise time standard such as UTC, the double-ended feature of the TRD approach provides the availability of precise time at each of the network nodes. This feature cannot be provided by any of the other techniques. The potential benefits for the synchronization subsystem are principally in the areas of compatibility and interoperability. There are, however, potential benefits completely divorced from the network timing function. These depend on the number of potential users of precise time that will be able to obtain it from the network nodes. We have not had the time to gauge this ancillary benefit.

Stability

Network stability in the control system sense is unconditionally assured in both the independent clocks and master-slave techniques. In the TRD approach, the only potential problem is bit errors in the information flow required for adaptive reorganization. Such errors could lead to portions of the network being configured as mutual synchronization subnetworks. This problem can, however, be avoided by the use of error correcting coding which essentially can reduce the probability of an error to an acceptable level. Because the required information rates are low, this is a very plausible solution.

As noted above, the stability of the mutual synchronization subsystem does present more stringent requirements. Stability can be ensured, however, by diligent loop gain control at each node.

Monitorability

The monitoring of the timing subsystem performance at a particular node can be accomplished equally well by any of the alternative techniques. The TRD technique, however, provides information not only of performance at the local node, but of performance at both the adjacent nodes and the network as a whole. What can be achieved is the identification of potential problems before they become serious problems. The information that is used is that present at the local node as a result of the double-ended feature, the adaptive reorganization feature, and the separation of error correction and error measurement feature.

Susceptibility

The independent clocks technique is the only technique that is not susceptible to jamming or electronic countermeasures. This is true since it is the only technique that does not depend on information transfer between nodes for operation.

The three disciplined techniques follow in increasing susceptibility as follows: TRD, master-slave, mutual-sync. This ordering is based on the degree to which the effects of a jammer, operating on a single node, will be felt throughout the network.

Flexibility

The only technique that may present an inherent problem with respect to growth and change in the network is the mutual-synchronization technique. Here, as discussed earlier, as nodes drop out or are added to the network, permanent changes occur in the network frequency.

Interoperability

Interoperability is concerned with the interfaces between the network and other communications systems. Considering the timing subsystem functions only, interoperability is enhanced if the phase/frequency offset between systems is minimized. This means less buffering and, therefore, less delay or concurrently lower probability of a bit slip for a fixed amount of buffering.

Following the same line of reasoning as that presented in the discussion on compatibility, the TRD approach offers potential benefits. Again, those potential benefits result only if a common timing reference exists for all networks.

Availability

Conclusions about availability of the timing function from technique to technique are difficult to quantify. This is true since availability can be built into any of the techniques by the addition of on-line sparing and greater redundancy.

However, it has been concluded that the hierarchical disciplined techniques do provide redundancy in highly connected networks. Also, it has been concluded that the TRD approach should have some availability benefits due to greater monitorability; and finally, that the independent clocks technique would require minimum sparing based on the lowest piece parts count.

Conclusions

The results of both the analysis and simulation efforts indicate that time reference distribution is the technique that provides the greatest overall utility. The TRD technique ranked higher in a majority of the desirable characteristics. The simulation indicated that TRD could maintain closer frequency coordination between nodal clocks in the face of both stress and normal operating conditions. The advantages of the TRD approach result from its key properties such as double endedness, adaptive reorganization, separation of error correction and error measurement, and provision of precise time. It is worth noting that all of these features require an overhead link capability, so there is a price to pay for this improved performance. In fact, the benefits of TRD that are highlighted at a highly connected, high-bit rate, "major node" are not as significant at a limited connected, low-bit rate, "minor node." Furthermore, at a minor node, the overhead channel required for TRD would represent a larger percentage of the data traffic. Therefore, the conclusion reached is that in a worldwide military digital network, the TRD technique is the appropriate timing alternative at the major nodes and that the master-slave technique (which does not provide the additional capability, does not require an overhead link, and is compatible with TRD) is the appropriate alternative at the minor nodes.

Acknowledgements

The authors are indebted to Dr. Harris Stover of the Defense Communications Agency for his guidance and helpful suggestions. The work on the computer simulation by Messrs. Ron Huhn and Carl Geisler is also gratefully acknowledged.

References

1. Darwin, G. P. and R. C. Prim, U.S. Patent No. 2,986,723 "Synchronization of a System of Inter-Connected Units."
2. Gray, D. A., "Transit-Time Variations in Line-of-Sight Tropospheric Propagation Paths," BSTJ, July, pp 1059-1068.
3. Lubowe, A. G., "Path Length Variation in a Synchronous Satellite Communications Link," BSTJ, Dec., 1968, pp 2139-2144.
4. Mensch, J. R., "Future DCS Objectives in Communication Network Timing and Synchronization," Proc. of Fifth Precise Time and Time Interval Planning Meeting, Dec. 1973.
5. Pan, J. W., "Synchronizing and Multiplexing in a Digital Communications Network," Proc. of IEEE Vol. 60, No. 5, May 1972, pp 594-601.
6. Pierce, J. R., "Synchronizing Digital Networks," BSTJ, March 1969, pp 615-636.
7. Stover, H. A., "Coordinated Universal Time (UTC) as a Timing Basis for Digital Communication Networks," EASCON Record, 1974.
8. Stover, H. A., "A Time Reference Distribution Concept for a Time Division Communication Network," Proc. of Fifth Precise Time and Time Interval Planning Meeting, Dec. 1973, pp 505-523.

*This work has been supported by the Defense Communications Agency under Contract No. DCA100-76-C-0028.

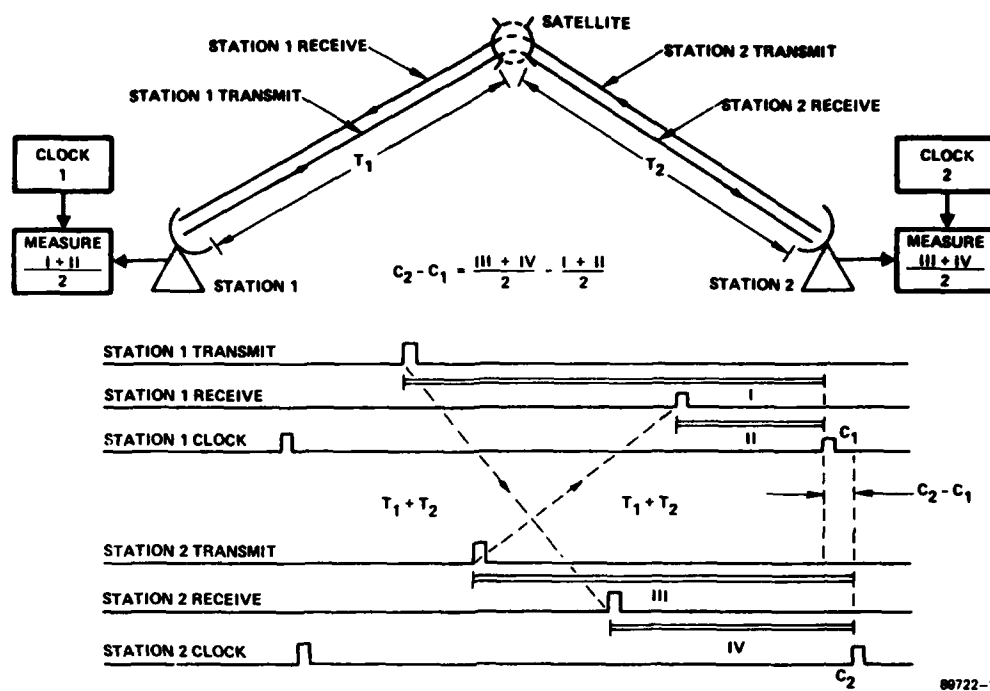


Figure 1. Time Transfer Timing Diagram for Satellite Link

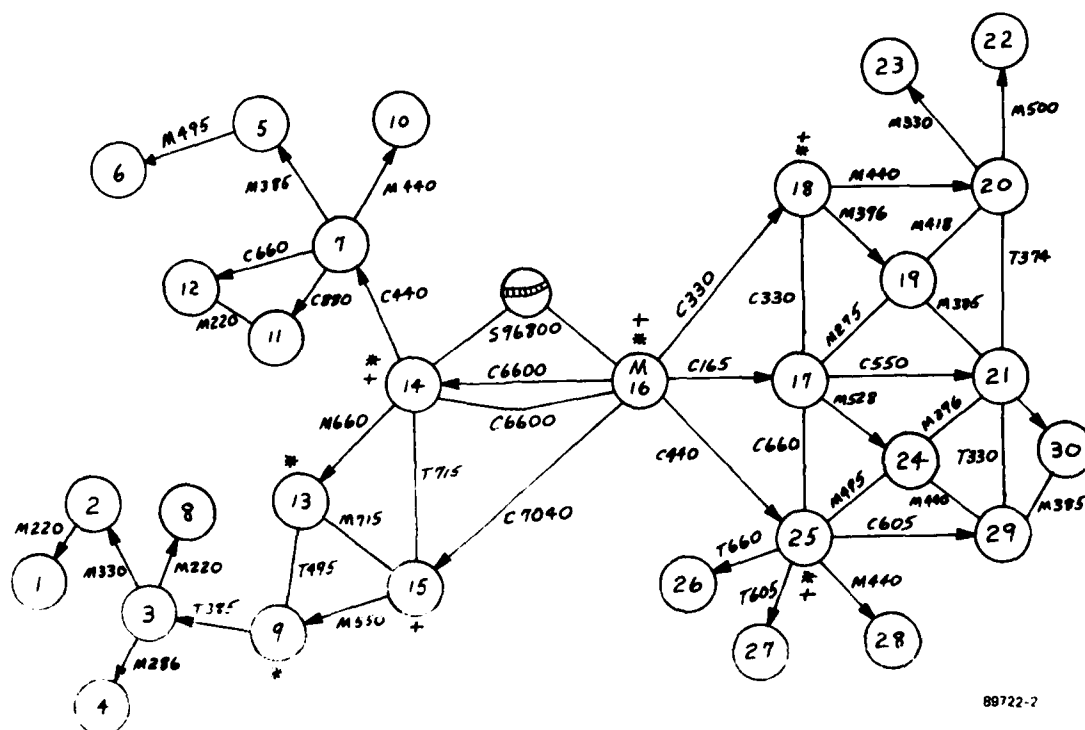


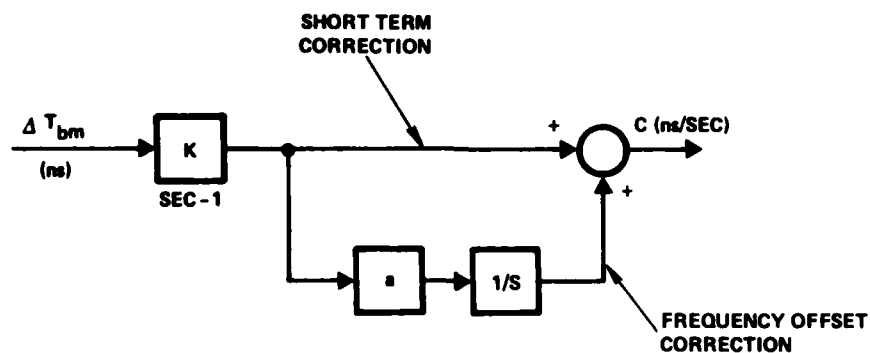
Figure 2. Dumbbell Node Configuration



Block diagram of a closed-loop system. The input is $\theta_i(s)$, which enters a summing junction with a positive sign. The output of the summing junction is $E(s)$, which enters a block labeled "LOOP FILTER" containing $F(s)$. The output of $F(s)$ enters a block labeled K_v . The output of K_v is $F_o(s)$, which is also the system output. A feedback path branches off from the output $F_o(s)$, passes through a block labeled $1/s$, and returns to the summing junction with a negative sign. The feedback signal is labeled $\theta_o(s)$.

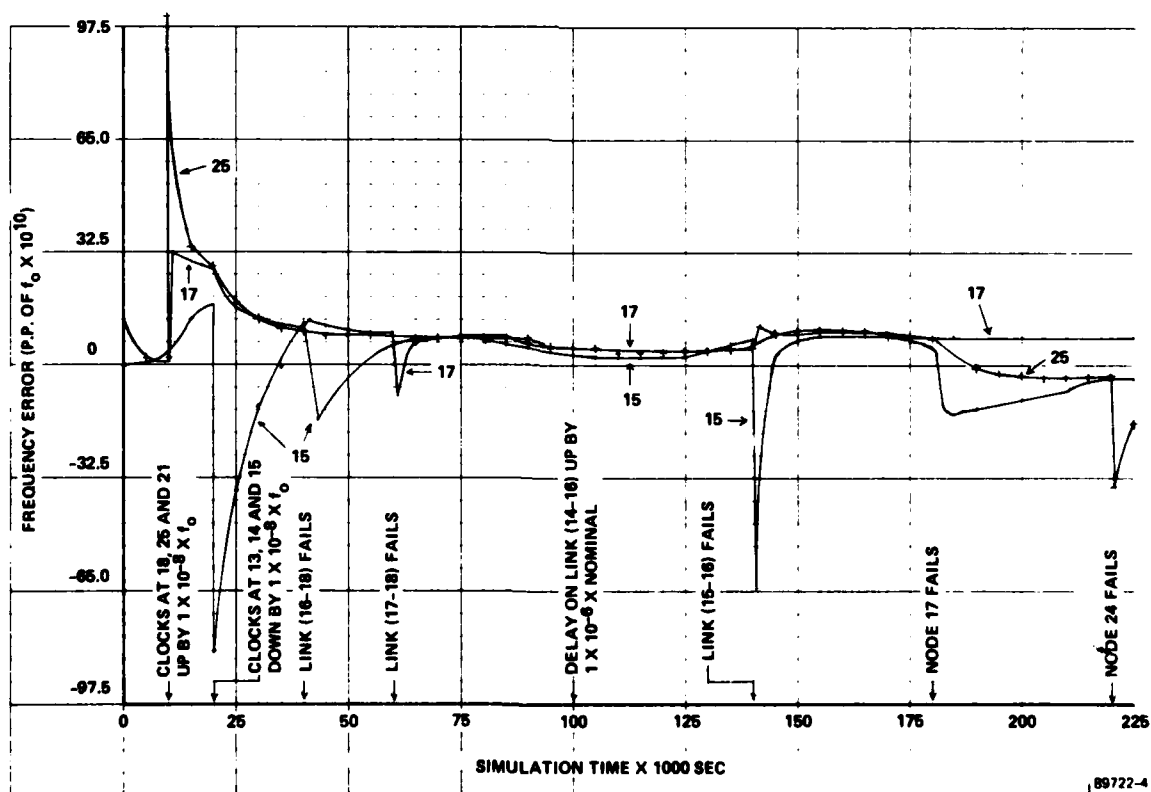
88807-18

445



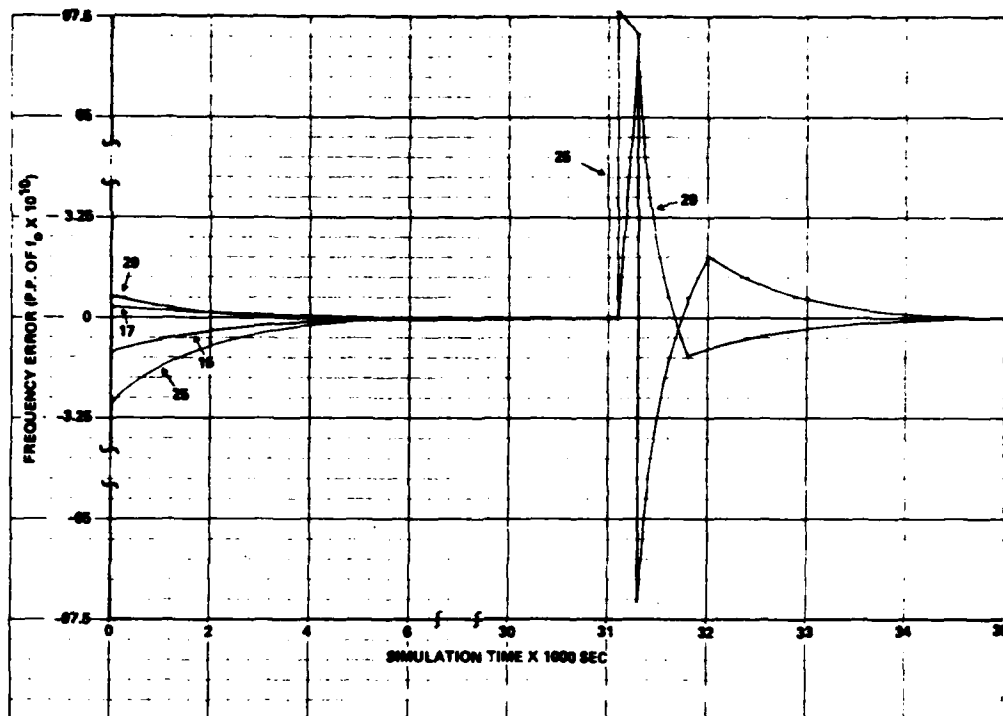
89643-2

Figure 5. Error Processing Filter for TRD



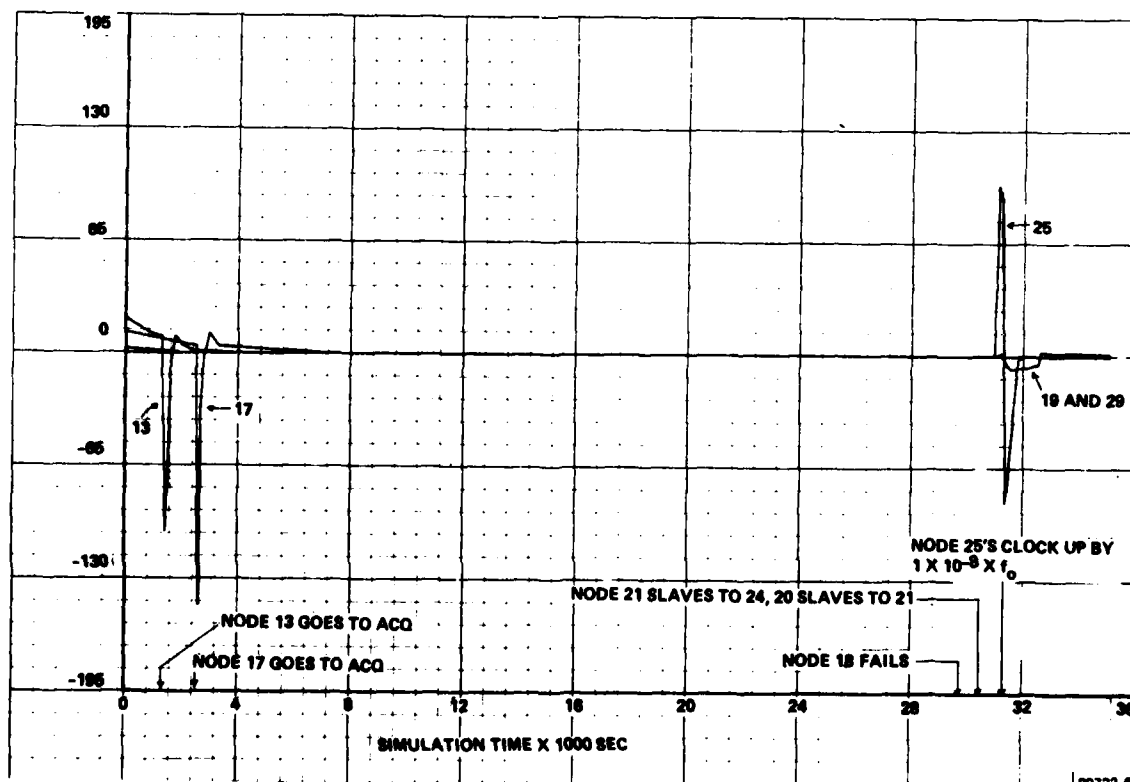
89722-4

Figure 6. Mutual Sync. Frequency Plots



89722-5

Figure 7. Master-Slave, Frequency Plots



89722-6

Figure 8. TRD Frequency Plots

FREQUENCY COORDINATION IN A NETWORK OF INTEGRATED SWITCHES

Henry A. Sunkenberg
Dr. Myron J. Ross

GTE Sylvania Incorporated
77 "A" Street
Needham Heights, Massachusetts 02194

1. Summary

This paper examines the synchronization requirements inherent in the control of switching and multiplexing functions in a network of digital time division switches interconnected by terrestrial transmission facilities. Specifically, two aspects of network synchronization will be described. The first relates to the problem of maintaining an accurate and stable clock at each switching node in the network so as to provide for the synchronous switching of digitized voice and data; the second is concerned with establishing and maintaining master frame synchronization in the digital trunk groups interconnecting the switches. Although we will primarily be concerned with a military environment in which both virtual circuit switching and packet switching are provided by an integrated switching system, many of the ideas expressed are applicable to both military and non-military networks providing only virtual circuit switched service.

2. Introduction

Digital transmission and switching techniques are now well established and promise to be more prevalent in the future. There are already either existing or planned commercial digital switching and long-haul transmission systems in the U.S., Canada, Japan, England and France.¹⁻⁵ In addition, the DoD is planning AUTOSEVOCOM II, a worldwide secure digital switching system which is scheduled for initial service around the 1980 time frame.⁶ An important characteristic of the above-mentioned systems is that they permit the sharing of transmission and multiplexing facilities by both digitized voice and data. It has been established that integrated transmission is more economical than separate transmission facilities. Recent studies^{7,8} have demonstrated that even greater economies can be obtained by integrating the switching function as well. Utilization of trunk groups is between 50 and 70 percent in a conventional virtual circuit switch engineered for 0.01 grade of service. Thus, approximately 30 to 50 percent of costly transmission capacity is unused in a conventional switch. However, when integrating both the circuit and packet switching functions, trunk utilization can exceed 90 percent. Therefore, considering the growth of data forecasted for the 1980's and beyond, it can be expected that totally integrated switching networks may be required as early as 1985.

To provide the necessary timing for such a network a synchronization scheme is required which meets the desired performance criteria, is cost-effective, and exhibits other key advantages such as reliability, survivability, modest complexity, maintainability, and interoperability with other systems. In this paper, both network timing and frame synchronization are examined with respect to these criteria.

3. Integrated Voice/Data Switching Concept

The network considered in this discussion consists of digital switches carrying voice and data integrated together in two classes: one class emulating circuit switching concepts and the other packet switching concepts. The master frame structure employed by these integrated digital switches (IDS) is shown in Figure 1.

As shown in this figure, the IDS concept^{8,9} is based upon the partitioning of transmission capacity into constant period self-synchronizing master frames. The self-synchronizing capability is realized by a start-of-frame marker (SOF) which indicates the beginning of each master frame. Following the marker, the remainder of the master frame is divided into two parts: a Class I region and a Class II region. The Class I region contains those types of traffic normally associated with circuit-switching (e.g., digitized voice, facsimile, and

low-speed video), while the Class II region incorporates traffic associated with packet or message switching (e.g., narrative record, interactive, query/response, data base update, and non-sensor bulk data).

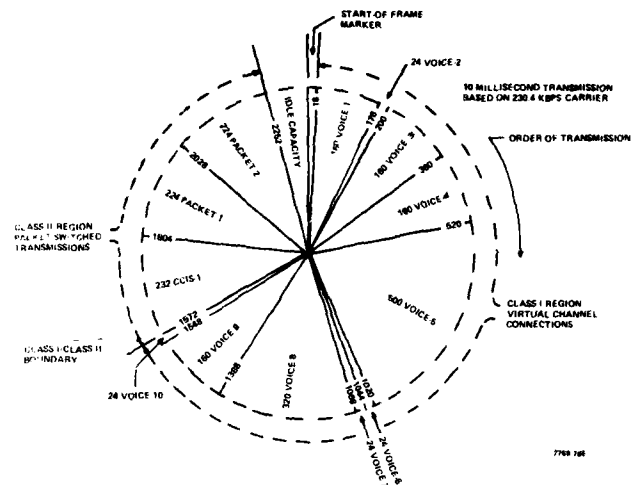


Figure 1. Integrated Voice/Data Master Frame Structure

In the Class I region, channel connections are allocated and maintained in accordance with link maps at each end of the link. Changes in these link maps are coordinated by common channel interswitch signaling (CCIS) messages. Class II data is handled as packets and is dynamically allocated as a function of the available space in each master frame. All CCIS messages required to initiate, allocate, coordinate, and terminate Class I calls are themselves treated as Class II data.

The typical IDS master frame shown in Figure 1 is based on a 10 msec period and a 230.4 kb/s transmission rate. Each of the wedges in the Class I region represents the bits associated with one call. For example, a 16 kb/s voice call is handled by reserving 160 bits in each master frame for the duration of the call. If the call lasts five minutes, then 160 bits are reserved for 30,000 master frames. The Class II region of packet-switched transmissions is shown in this specific case carrying one CCIS message and two data packets. These could be of different sizes and ordered by precedence. The end of a master frame is marked by the SOF marker of the next subsequent master frame.

Within the Class I region, call location reflects primarily the age of the call. The closer a given call is found to the SOF marker, the older in general that call is. As calls are terminated, the Class I region is compacted and everything moves up, including the Class I/Class II boundary. This shrinking of the Class I region causes the Class II region to expand, thereby allowing more data packets to be transmitted. In this way, the regions react dynamically to changes in one another, resulting in a maximization of the throughput.

4. Network Synchronization

The problem being considered is to define the means by which the IDS network synchronizes switching and multiplexing functions so that all data which is received at a node from a connected node and is to be

transmitted to a connected node, is delivered to the proper destination, in the proper sequence, and without the addition of extraneous bits. There are two aspects to the overall network synchronization problem, internode and intranode synchronization. The former includes network timing and master frame synchronization on a link-by-link basis. The latter concerns master frame alignment and coordination at each node. Maintenance of network timing in the IDS network constitutes the same problem it does in any digital network and is discussed in Section 4.2. Frame Synchronization is examined in Section 4.3. Frame alignment and coordination for IDSs is somewhat more complicated than in a conventional TDM/digital switch due to the dynamic nature of the IDS master frame. This problem is considered in Section 4.4. Throughout this discussion, it will be assumed that all links connecting nodes transmit at the same data rate and possess the same master frame period.

Before examining the two aspects of network synchronization discussed above, we will consider the two basic problems with which all synchronization plans must contend: nodal clock frequency drift and transmission delay variations.

4.1 Frequency Drift and Transmission Delay

Two types of clocks are primarily used in digital networks, crystal clocks and atomic clocks. Figure 2 illustrates typical characteristics for commercially available clocks.^{10,11} $\sigma_y(\tau)$, the square root of the Allan Variance¹², provides an invariant measure of the short-term stability of an oscillator. It also characterizes the long-term stability as long as $y(t)$ is assumed stationary over the longer averaging period. Here, we assume that $\sigma_y^2(\tau)$ is valid at least for τ less than or equal to a year.

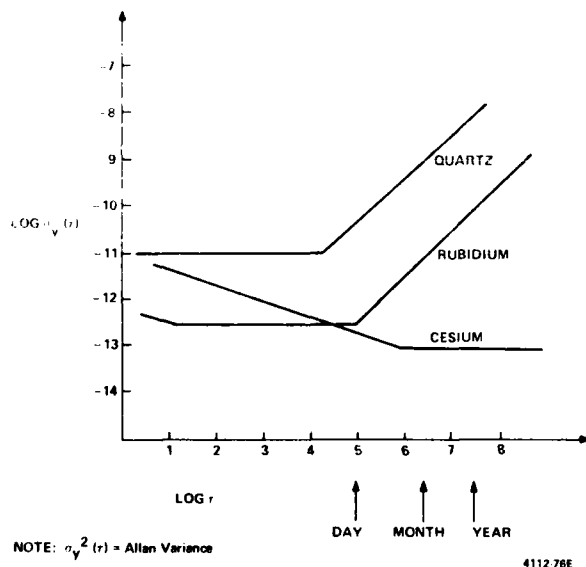


Figure 2. Typical High Precision Clock Performance

For the purposes of network timing, it is sufficient to consider only the long-term stability of the clocks. Thus, the stability of a commercial crystal clock may vary, depending on cost, between 10^{-6} and 10^{-10} ; the stability of a commercial atomic clock is on the order of 10^{-12} . Because of the greater stability of atomic clocks, they are usually used as the master clock(s) within a network. Crystal clocks are usually used for backup purposes or as nodal clocks (possibly phase locked to a master) at lower level nodes. Because all clocks exhibit frequency drift, network timing is usually recalibrated once or twice a year against a primary standard.

Transmission delay variations also cause timing problems; however, their effects manifest differently than do the effects of nodal clock variations. To see this, consider two interconnected and independently timed nodes A and B. If the transmission delay between A and B increases (decreases) in both directions, then more (less) bits are stored in

the link, and the receiver buffers at each end of the link will empty (fill) simultaneously. On the other hand, if the nodal clock at A is faster (slower) than the nodal clock at B, then the buffer at A will empty (fill), while the buffer at B will fill (empty). It should be noted that by comparing buffer changes at both ends of a link, it should be possible to determine whether buffer changes are due to clock or transmission delay variations.

Estimates¹³ of transmission delay variations for coaxial cable and microwave radio are as follows:

a. Coaxial Cable

1. The dominant cause of delay variations is linear expansion caused by temperature change.
2. Changes in transmission delay occur slowly.
3. For a path length of 3000 miles and a transmission rate of 1.544 Mb/s, the change in the number of bits stored in a cable due to a 22°C temperature change is approximately 10 bits.

b. Microwave Radio

1. Delay variations are due to changes in temperature, pressure, humidity and probably rain.
2. Rapid and long-term delay variations are possible.
3. Average variations in link bit storage for a 1.544 Mb/s transmission rate are (1) daily, 0.49 bits; (2) monthly, 2.5 bits; (3) Yearly, 7.4 bits.

4.2 Network Timing

It is not possible to have all nodes within a network in perfect synchronism. Consequently, any data received at a node from a connected node must be brought into synchronism with the local clock. There are two primary methods for performing this internode synchronization:

- a. A synchronous or clocked approach where each nodal clock is controlled so that all clocks maintain the same average frequency.
- b. An asynchronous or unclocked approach where all nodal clocks are independent, extremely stable and free running, and buffers are used to absorb frequency errors.

The four major internode synchronization methods^{14,15} are:

- | | |
|------------------------|-----------------------|
| a. Independent Clocks | } Asynchronous Clocks |
| b. Bit Stuffing | |
| c. Master/Slave | } Synchronous Clocks |
| d. Frequency Averaging | |

Included within these four methods are numerous variations such as the modified (or hierarchical) master/slave, external time reference, and the time reference distribution systems. In general, the choice of an 'optimum' scheme would depend on evaluation criteria (network objectives) and the network topology being considered. Table 1 illustrates the authors' attempt to objectively rate the major network timing schemes with respect to five categories. Implementation cost denotes both hardware and software development costs. This criterion is used because many timing schemes require software for the nodal timing unit controller and because the basic hardware costs do not vary considerably between the alternatives relative to total network costs. Reliability/availability includes both equipment failure rates and the impact of the environment; for example, bit stuffing codeword errors can result in loss of bit count integrity (slips) in the affected channels. Survivability reflects the impact of the loss of nodes or links. Both the reliability and survivability categories in Table 1 do not consider the inclusion of backup systems. Monitoring/maintainability considers the complexity of the hardware and the performance assessment information that can be derived from the timing system, while service continuity reflects the frequency with which network users can expect slips.

Thus for any given application, the network designer would have to weigh each of the five categories to reflect his requirements, and possibly consider other criteria should these be insufficient.

Table 1. Relative Comparison of Bit Synchronization Schemes

	IMPLEMENTATION COST	RELIABILITY/AVAILABILITY	SURVIVABILITY	MONITORING/MAINTAINABILITY	INTEROPERABILITY	SERVICE CONTINUITY
INDEPENDENT CLOCKS	2	2	1	2	2	3
BIT STUFFING	1	3	2	3	1	2
MASTER/SERVE	1	1	3	2	2	1
MODIFIED MASTER-SLAVE	2	2	2	1	2	1
FREQUENCY AVERAGING	2	2	2	1	2	1

The IDS is being considered for application within the future military environment. Within this framework, the network model used for evaluation of the IDS is a two-level hierarchical configuration with ten tandem nodes, highly interconnected, and fifty regional or access nodes (five per tandem node), each feeding one tandem node. This model is derived from the current thinking of the Defense Communications Agency with respect to the evolution of the Defense Communications System¹⁶ toward a structure which provides integrated voice and data in a common-user network. Thus, network timing in the IDS network could be accomplished by any of the major timing schemes with the probable exception of bit stuffing; this is due to the variable channel rates being processed and the fact that the switching is performed by software.

Based on an analysis of the requirements of the military communications and the postulated network structure with respect to Table 1, the modified master/slave technique was selected as the preferred scheme for use with IDS switches. In this approach, each tandem node in the network would serve as a local master for all regional nodes (slaves) to which it is directly connected; thus, each local master and its connected slaves would form, in effect, an independent subnetwork. For the particular configuration described (i.e., 10 tandem nodes and 50 regional nodes), the network would comprise 10 subnetworks with each subnetwork containing 5 slave nodes. A typical network configuration is illustrated in Figure 3. Observe that all subnetworks are interconnected at both the local master and slave levels. It is assumed that each subnetwork would be defined by geographical as well as traffic considerations.

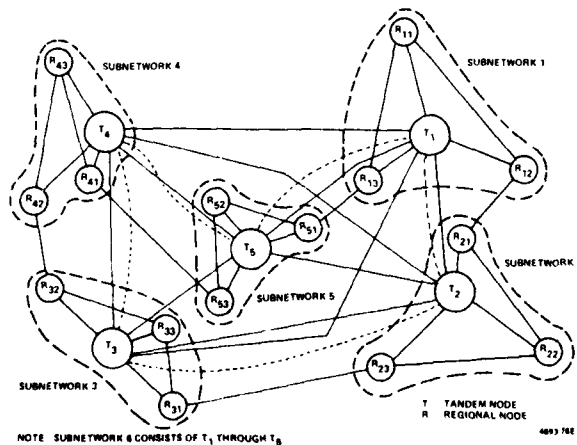


Figure 3. Typical Network Configuration

There are three alternatives for synchronizing the local masters: frequency averaging, a master/slave arrangement, or independent atomic clocks. The first two approaches would result in an overall synchronous network. The independent clock approach would result in asynchronous operation between subnetworks and as previously mentioned, synchronous operation within a subnetwork. Again, consideration of the assumed network structure, specifically the fact that the tandem nodes are fully interconnected, reduces the selection to frequency averaging and the master/slave approach. This follows, since for the given network configuration independent atomic clocks

would not be significantly more survivable than either of the other two approaches, and would definitely be higher priced. Note that even if the tandem nodes are not fully interconnected in an actual implementation, the degree of connectivity is likely to be sufficiently high to make the elimination of the independent clock approach justifiable. An additional consideration is the fact that future communications will likely see a greater and greater use of end-to-end encryption, making asynchronous techniques less and less attractive. The choice between the master/slave approach and frequency averaging is not so clear. It would appear, however, that the advanced state-of-the-art in master/slave techniques and the numerous stability problems inherent in frequency averaging makes the former the practical approach. For this reason, the master/slave technique is the approach that was selected to synchronize master nodes in the IDS design.

The operation of the chosen timing scheme will be illustrated with respect to Figure 3. In accordance with the chosen synchronization plan, a particular tandem node is designated master node for subnetwork 6. That is, during normal operation, all other tandem nodes slave their timing to this node. Because the tandem nodes are fully interconnected, any tandem node could be chosen for this purpose. It is assumed for illustrative purposes that T5 (in Figure 3) is designated master. It should be realized though that T5 serves as master not only for the other tandem nodes/local masters but also for all regional nodes (slaves) via the three level hierarchy which results from the modified master/slave implementation. As a practical means of implementing this master/standby master arrangement, it will simply be necessary to assign a slaving precedence to the tandem nodes. The master would be assigned the highest precedence n (where n is the number of tandem nodes) and the standby masters successively lower precedences (n-1, n-2, n-3, ..., 1) as determined in an appropriate manner. Then, at any tandem node which is slaving its timing, the nodal clock would slave to the link which has the highest precedence level and from which proper timing signals are being received.

As presently described, the network is extremely vulnerable to loss of the master. In fact, in order to avoid frequent link buffer overflow/underflow in this event, it would be necessary to equip all nodes with either very stable clocks, large link buffers, or an appropriate combination of the two. As a means of avoiding this necessity and to greatly improve survivability, at least one other tandem node will be capable of assuming the master role. It would be possible to have all tandems so equipped; however, the actual number of standby masters used would be determined from the degree of survivability desired, the accuracy of the nodal clocks, and the size of the link buffers. It is also possible for only part of the timing tree to reconfigure. For example, if the link carrying timing signals from T5 to T2 were to fail, then T2 would automatically slave to T1, assuming it is the node with the second highest precedence level. Timing in subnetworks 1, 3, 4 and 5 would be unaffected by this partial reconfiguration.

In the IDS network, status information concerning the various nodal timing supplies would be exchanged between switches in the form of CCIS messages. Additionally, the precedences of the tandem nodes could be periodically changed for increased security and survivability. Again, CCIS messages could be used to accomplish this since, in the military environment, all interswitch trunk groups will be bulk encrypted.

Figure 4 is an example of the design of the timing unit which could be employed at a tandem node. The slave node timing unit is

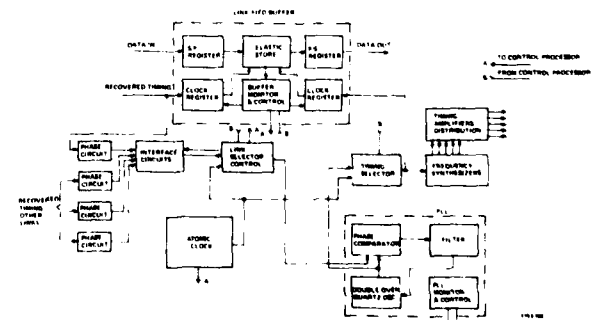


Figure 4. IDS Timing Unit - Master and Standby Master Configuration

essentially equivalent except that it uses two phase locked loops instead of the atomic clock and single phase locked loop used at master/stand-by master locations. Redundant clocks at each node and multiple atomic clocks are specified to meet the stringent availability/survivability requirements inherent in a military environment.

In order to provide for communications within or between sub-networks which for one reason or another are unable to operate in a slaved node, all nodal clocks will be capable of free-running. The required stability of the nodal clocks in the free-running mode is a function of the size of the link buffers. During normal operations, the link buffers at each node should maintain a long-term average fill of one-half. Short-term variations from this value are to be expected due primarily to jitter introduced by repeaters, transmission delay variations, and tracking inaccuracies in the phase locked loops. Major excursions from the half-way point would arise when a nodal clock is free-running or the phase locked loop is acquiring lock following an outage or possibly a switch from the network master to a standby master. The link buffers must be sufficiently large to absorb the maximum bit slippage to be expected during the two types of events. The more severe of the two events is the case of the free-running oscillator and the link buffers will be designed for this eventuality.

Based on the typical stabilities of commercially available clocks, as shown in Figure 2, the buffering required to compensate for nodal clock drift in the free running mode is shown in Figure 5. For the IDS design, a buffer length of 256 bits was chosen. Including the effects of transmission delay variations as described in Section 4.1, this permits a free running capability at all nodes of approximately a day.

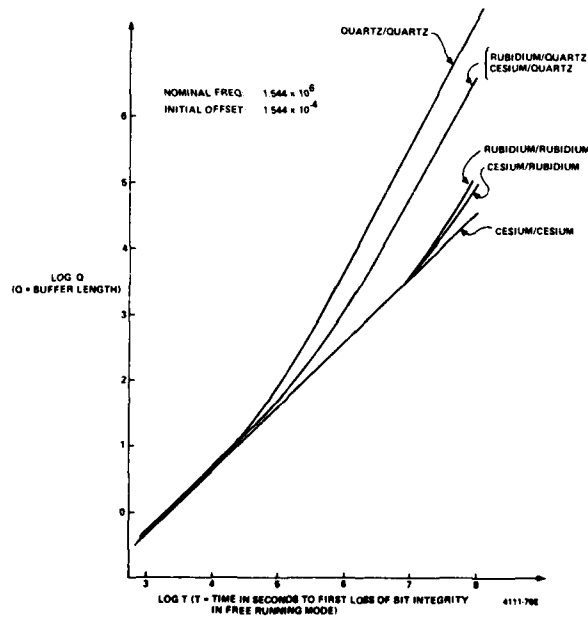


Figure 5. DAX Buffer Requirements

4.3 Master Frame Synchronization

In each master frame, the location of each virtual circuit switched connection is only known relative to the starting position of the master frame. Therefore, before Class I information can be transmitted over an inter-IDS link, master frame synchronization must be established and maintained. The method by which the IDS accomplishes this type of synchronization is the subject of this section. It is assumed that network timing as described in the previous section has already been achieved.

Those portions of a IDS concerned with frame synchronization are shown in Figure 6. The frame maintenance unit (FMU) is the hardware which realizes the chosen synchronization procedures under software control. An FMU is associated with each full-duplex inter-IDS link and performs two basic functions. It maintains synchronization in the receive direction and assists its companion FMU at the remote switch in maintaining synchronization in its receive direction.

In order to accomplish these functions, three start-of-frame markers (SOF) are used:

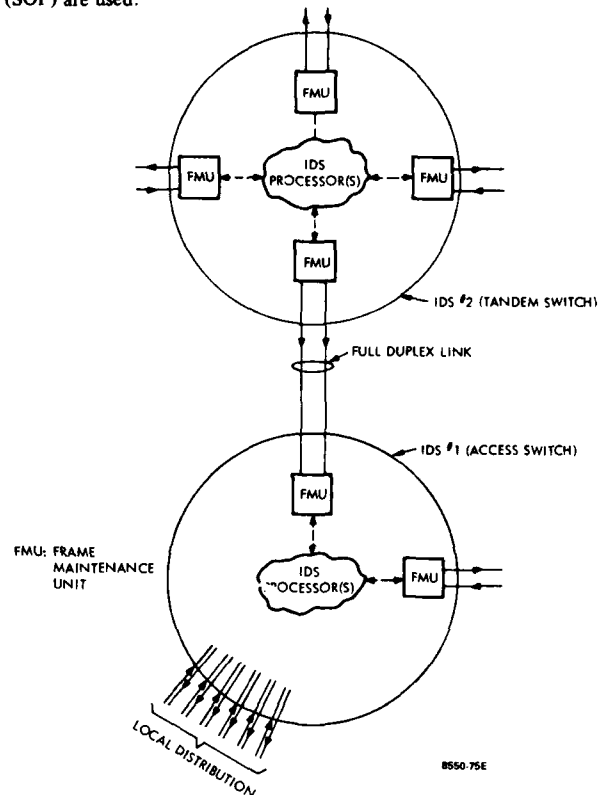


Figure 6. Frame Synchronization Equipment

- SOF-1: In-sync pattern. When transmitted from IDS 1 to IDS 2 (see Figure 6), it provides a synchronization pattern for IDS 2 and informs IDS 2 that IDS 1 is in-sync in its receive direction.
- SOF-2: Request-for-frame pattern. When transmitted from IDS 1 to IDS 2, it provides a synchronization pattern for IDS 2 and informs IDS 2 that IDS 1 is out-of-sync in its received direction.
- SOF-3: Out-of-sync pattern. When transmitted from IDS 1 to IDS 2, it permits IDS 2 to resynchronize within a small number of frames.

At this point, it is tentatively assumed that SOF-1 and SOF-2 are short sequences on the order of 16 bits and that SOF-3 is a much longer sequence. The exact length of the SOFs required to achieve adequate performance is determined subsequently.

The FMU has two operational states or modes: a frame maintenance mode and a frame acquisition mode. When operating in the frame maintenance mode, the FMU assumes that the link is in-sync and monitors for loss of this condition. To ensure that the FMU performs satisfactorily in this mode, the algorithm by which it chooses between in-sync and out-of-sync must be such as to simultaneously minimize the following probabilities.

- a. Prob (false out-of-sync alarm), P_{FOA} : this denotes the probability of the FMU indicating out-of-sync when in fact the link is in-sync
- b. Prob (loss-of-sync miss), P_{LM} : this denotes the probability of the FMU indicating in-sync when in fact the link is out-of-sync.

In the frame acquisition mode, the FMU assumes an out-of-sync condition and scans the input for a valid SOF. This portion of the acquisition mode is referred to as the search phase. Once the FMU locates what is believed to be the valid synchronization position, it

terminates the search phase and enters the check phase. During this next phase, the FMU attempts to verify that the assumed synchronization position is in fact the true synchronization position. If the check phase indicates that the assumed position is incorrect, the FMU returns to the search phase; otherwise, the FMU assumes that resynchronization has been achieved and returns to the maintenance mode. A measure of the performance of the FMU in the acquisition mode is the Prob (false in-sync alarm) of P_{FIA} . This is the probability that acquisition mode locks on to a position which is not the true synchronization position. The acquisition mode must be designed to make this probability negligible. Another factor of interest in the acquisition mode is the average time to acquire synchronization, $\langle u \rangle$. This quantity denotes the average number of bits required to lock on to the correct synchronization position and excludes the check phase when the correct synchronization position is found. Since on the average, the FMU must check half the frame before reaching the transmitted SOF, $\langle u \rangle$ is lower bounded by one-half frame.

4.3.1 SOF Bit Pattern Considerations - A valid SOF can appear in any of three regions within the master frame: the true synchronization position; a region comprised partly of the true synchronization position and partly of random data, denoted as the overlap region; and the pure-random data region. In the frame maintenance mode, the FMU scans only region 1 while testing for valid SOFs. Clearly, then, for a given SOF length, the probability of obtaining a false synchronization pattern correlation is independent of the particular bit pattern used and depends only on the error environment. In the search phase, the FMU sequentially scans all these regions searching for a valid SOF. For regions 1 and 3, the same conclusion applies; namely, for a given SOF length, the bit pattern used plays no role in determining the probability of obtaining a false synchronization pattern correlation. However, the SOF bit pattern does impact on the probability of a false correlation in the overlap region since this probability is a function of the number of bits in agreement with the true SOF. To minimize the impact of SOF simulations in region 2, it is necessary to satisfy the following bit pattern restriction,¹⁷

$$(\rho_0, \rho_1, \rho_2, \dots, \rho_{j-1}) \neq (\rho_{N_1-j}, \rho_{N_1-j+1}, \dots, \rho_{N_1-1})$$

$$j = 1, 2, \dots, N_i$$

where $(\rho_0, \rho_1, \dots, \rho_{N_i-1})$ = SOF-*i* and N_i = length (SOF-*i*).

This ensures that in an error free environment there can be no simulation of the SOF in region 2 and also serves to minimize the probability of this event in a noisy environment. As a consequence of the above SOF bit pattern restriction and the fact that the acquisition mode includes a check phase, it is assumed that P_{FIA} is calculable by considering only region 3.

Due to the assumed length of SOF-3, it is not necessary that it strictly adhere to the above bit pattern restriction. In fact, as long as it is of reasonable length (e.g., > 25 bits), SOF-3 can be chosen for convenience rather than as a result of careful analysis. Thus, if SOF-2 is chosen to be the binary complement of SOF-1 (thereby maximizing the Hamming distance between the two), then SOF-3 can be the convenient pattern SOF-1 || SOF-1 || SOF-2. Such a choice for SOF-3 would simplify the hardware implementation of the FMU and the transition of the FMU from the acquisition mode to the maintenance mode.

4.3.2 Frame Maintenance Mode Performance - The FMU enters the frame maintenance mode upon acquisition of frame synchronization and maintains this mode until it determines that synchronization has been lost. There are numerous methods by which the FMU can decide between an in-sync and out-of-sync condition. The method to be used here is an *n*-frame test using non-overlapping frames and a fixed threshold decision process. The primary reasons for selecting this method are as follows:

- The implementation is straightforward and is easily realized with hardware or software
- The implementation provides excellent performance for carefully chosen values of *n*, the fixed threshold, and other test parameters.

The specific operations of the chosen technique require that for

each master frame for *n* successive master frames, the FMU correlate (compares bit-by-bit) the bit pattern in the assumed synchronization position with the expected SOF. If a correlation results in *r* or fewer bit disagreements, a positive correlation is recorded; otherwise, a negative correlation is recorded. Whereupon, if in *n* such comparisons the number of positive correlations equals or exceeds some fixed threshold, *t*, an in-sync condition is declared and the test is repeated with the next *n* master frames. If the number is less than this threshold an out-of-sync condition is declared, the frame maintenance mode is terminated, and the FMU initiates the frame acquisition mode.

The performance of the frame maintenance mode is measured in terms of P_{FOA} and P_{LM} . These probabilities, as may be expected, are intimately related to *n*, N_i , and *t*. Unfortunately though, P_{FOA} and P_{LM} have opposing requirements, that is, as the parameters are varied to decrease (increase) P_{FOA} , P_{LM} increases (decreases). A full parametric analysis to jointly optimize P_{FOA} and P_{LM} would require extensive computer time. In order to avoid this necessity, a value of *n* will be chosen which results in a reasonable mean time to determine loss of synchronization. For this value of *n*, the test threshold will be parametrically varied for likely N_i candidates to determine if reasonable values of P_{FOA} and P_{LM} result. If not, the process will be repeated for different values of *n*. Although this procedure may not result in an optimal solution, it does provide excellent results. As a starting point, the following is assumed, *n* = 10 master frames, N_i = 13 or 16 bits and $\tau = 1$. Additionally, a bit error rate, *e*, of .01 is used in the analysis. Whereupon, P_{FOA} , in terms of previously defined parameters, is given by:

$$P_{FOA} = \sum_{k=n-t+1}^n \binom{n}{k} (P(\bar{A}|B))^k (P(A|B))^{n-k}$$

where

$$P(A|B) = \sum_{k=0}^{\tau} \binom{N_i}{k} e^k (1-e)^{N_i-k} \quad \text{and}$$

$$P(\bar{A}|B) = 1 - P(A|B).$$

A determination of P_{LM} depends on what region of the master frame the FMU is monitoring. As discussed previously, a worst case assumption is to assume that the FMU is monitoring the random data region. In which case

$$P_{LM} \approx \sum_{k=t}^n \binom{n}{k} (P(\bar{A}|B))^k (1 - P(\bar{A}|B))^{n-k}$$

where

$$P(\bar{A}|B) = \sum_{k=0}^{\tau} \binom{N_i}{k} 2^{-N_i}$$

Figure 7 illustrates the result of the P_{FOA} and P_{LM} calculations for the specified parameter values. As may be seen from this figure, the chosen parameter values provide excellent results; however, a decision as to which SOF length is preferable will be deferred until the impact of frame acquisition has been considered.

4.3.3 Frame Acquisition Mode Performance - The operation of the FMU during frame acquisition is a two-phase process. It should be noted that the FMU must be capable of synchronizing with SOF-2 or SOF-3. Since SOF-2 is the shorter of two sequences, the frame acquisition mode must be designed to perform adequately with SOF-2. However, SOF-2 is also used in the frame maintenance mode; thus, its length must be such as to satisfy the requirements of both modes.

The principal parameters which control the performance of the frame acquisition mode are *r*, *s* (the length of the check phase) and N_i , where performance is measured in terms of $\langle u \rangle$ and P_{FIA} . It can be shown¹⁸ that subject to assumptions which are met here, the average time to acquire synchronization in the acquisition mode is

$$\langle u \rangle = \frac{M+1}{2} + M \frac{P(\bar{A}|B)}{P(A|B)} + \frac{M(M-1)}{2} P(\bar{A}|B) (1 + 2 \frac{P(\bar{A}|B)}{P(A|B)}) s$$

where $P(A|B)$, $P(A|\bar{B})$ and $P(\bar{A}|B)$ are as defined in the previous section, and M is the length of the master frame (assumed here to be 15440 bits).

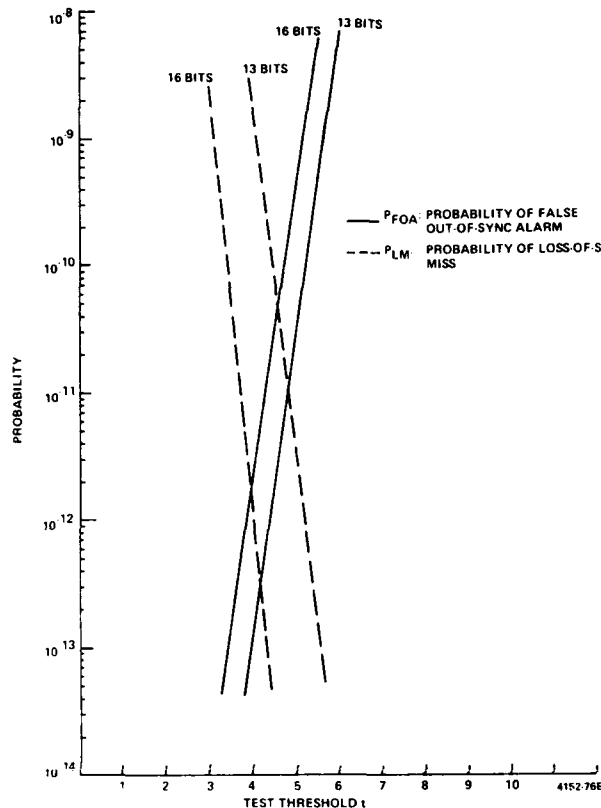


Figure 7. P_{FOA} and P_{LM} as a Function of Test Threshold With N_1 as a Parameter

Observe that if in the search phase the probability of accepting a false synchronization position ($P(A|\bar{B})$) is large, then $\langle u \rangle$ is proportional to the square of the frame length. For this reason it is important to minimize $P(A|\bar{B})$. A related parameter of interest is the variance of u . However, this is an extremely difficult quantity to calculate. In its place, the parameter P_{FI} will be calculated. This is the probability that the first indication of frame synchronization in the search phase is the true synchronization position. This is given by¹⁹

$$P_{FI} = \frac{1}{M} \frac{P(A|B)}{P(A|\bar{B})} \frac{1 - (1 - P(\bar{A}|\bar{B}))^M}{1 - P(\bar{A}|B) - (1 - P(\bar{A}|\bar{B}))^M - 1}$$

where $P(\bar{A}|\bar{B}) = 1 - P(A|\bar{B})$

The remaining performance indicator of interest is P_{FIA} . For reasons discussed previously, this probability will be derived by neglecting the contribution of the overlap region of the master frame. It can be shown¹⁷ that the probability of choosing a simulated SOF in the random data region instead of the true synchronization position (P_{RS}) is given by

$$P_{RS} = \sum_{k=1}^a (-1)^{k+1} \frac{1}{k+1} \binom{M-N_1 - (N_1-1)k}{k} 2^{-N_1 k}$$

where a is the largest integer smaller than or equal to $(M-N_1)/N_1$ and where it has been assumed that $\tau = 0$ (based on the mean number of bit errors expected). Therefore, if the check phase requires s master frames for verification of synchronization then P_{FIA} is approximately given by

$$P_{FIA} \approx 2^{-N_1 s} P_{RS}$$

Table 2 provides an evaluation of the performance of the acquisition mode for different parameters of interest. As may be seen there, in order to obtain tolerable values for P_{FIA} when synchronizing with SOF-2, two frames are required in the check phase for both the 13- and 16-bit lengths. However, a single check frame is adequate when synchronizing with N_3 .

Table 2. Performance of the Frame Acquisition Mode

SOF	P(A B)	P(A B)	P(A B)	s = 1		s = 2			
				$\frac{P(A)}{P(B)}$ [frames]	P	$\frac{P(A)}{P(B)}$ [frames]	P		
N ₂ = 13 bits	.878	.122	1.22x10 ⁻⁴	1.84	.274	6.65x10 ⁻⁵	3.05	274	8.11x10 ⁻⁹
N ₂ = 16 bits	.851	.149	1.53x10 ⁻⁵	.83	.961	1.66x10 ⁻⁶	.99	.961	2.5x10 ⁻¹¹
N ₃ = 19 bits	.676	.324	1.82x10 ⁻¹²	.98	~1	2.54x10 ⁻²⁰	.98	~1	4.62x10 ⁻³²
N ₃ = 48 bits	.617	.383	3.55x10 ⁻¹⁵	1.12	~1	9.74x10 ⁻²⁶	1.12	~1	3.46x10 ⁻⁶³

$a = .01$

$\tau = 0$

3879 JPL

At this point it will be possible to select a length for SOF-1 which, in turn, determines N_2 and N_3 . As a result of the frame maintenance analysis, it was seen that both the 13- and 16-bit lengths provide excellent performance. However, with regard to frame acquisition, the 16-bit length decisively outperforms the 13-bit length. As a result, a length of 16 bits is proposed for N_1 . A choice of a particular bit pattern for SOF-1 is not critical but must be consistent with results of Section 4.3.1.

4.4 Master Frame Coordination

At any IDS, the start-of-frame markers in all outgoing links are in time alignment and the start-of-frame markers in all incoming links are randomly aligned with respect to outgoing links. This is the same timing relationship that exists in a conventional TDM system. Unlike the conventional TDM, there is no absolute need to align master frames in the IDS since the Class I region map which exists at each IDS for each terminating link uniquely defines the location of each channel within its master frame. In effect, the CCIS messages provide the control channel which is required when frame alignment is not used.

Related to frame alignment is the timing or coordination problem generated by the reassignment of channels within a master frame. To illustrate the nature of this problem, consider the procedure required to establish a call at an IDS. Referring to Figure 8a, a Class I call is allocated channel 6 on the incoming link and channel 5 on the outgoing link. Note that the master frames on the incoming and outgoing links are out of alignment. Data bits for the call first appear in incoming master frame No. 1 and are placed in outgoing master frame No. 1. They could have been placed in outgoing master frame No. 2 but this would have added additional cross-office delay, a situation which, if possible, should be avoided. Similarly, data bits received in incoming master frame No. 2 are placed in outgoing master frame No. 2; this procedure continues frame-by-frame until the conclusion of the call.

The problem with this particular channel assignment is that if the first four calls in the outgoing link were to terminate while all calls in the incoming link were to continue, channel 5 would be moved to a position adjacent to the start-of-frame marker; consequently, in the frame in which this move were effectuated, the call bits to be placed in channel 5 in the outgoing link would be required before they were received on the incoming link. The probable reaction of the ISN to this situation would be to repeat the previous frames call bits, as shown in Figure 8b. This channel slip would appear to the customer as a repetition or 'hiccup' of data, a channel period in length.

There are two ways to avoid this 'hiccup' effect. One is to delay the call bits from each incoming channel one master frame period before placing them in an outgoing channel; the other is to never place the call bits from an incoming channel in an outgoing channel

whose corresponding start-of-frame marker begins prior to the end of that incoming channel. Both methods assure that at any tandem node an outgoing channel can never be moved forward to the extent that it slips by (precedes in time) its corresponding input channel.

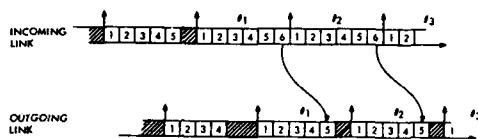


Figure 8a. Channel Allocation at a Tandem Mode

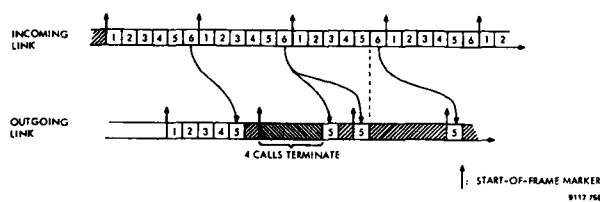


Figure 8b. Example of a Channel Slip

5. Conclusions

In this paper, network timing and frame synchronization have been examined for application to a switching network whose requirements are based on projections of the future Defense Communication System. This network consists of 60 nodes, including 10 tandem and 50 regional nodes, arranged in a two-level hierarchy, and carries voice and data integrated together in two classes: one class emulating circuit switching concepts and the other packet switching concepts. For this network structure, the modified master/slave approach employing multiple masters was selected as the preferred implementation scheme for network timing. The rationale for this selection is based on its high performance with respect to the evaluation factors implementation cost, reliability/maintainability, survivability, monitoring/maintainability, interoperability, and service continuity. A double-ended frame synchronization plan was chosen as the method of establishing and maintaining frame synchronization on the link's interconnecting switches. This plan utilizes three start-of-frame markers to keep each node at each end of a link aware of the synchronization status at the other end. Included were analysis of synchronization acquisition and maintenance and recovery procedures in the event of loss-of-synchronization. A timing or coordination problem for the virtual circuit-switched channels in transferring data from incoming to outgoing links, due to the dynamic nature of the integrated master frame structure, was also illustrated.

References

1. "Digital Data System," BSTJ, Vol. 54, No. 5, May-June 1975.
2. Verma, P., et.al., "Operating Experience with the Canadian Dataroute Synchronization System," ICC-1976, pp. 25/16-20.
3. Inoue, N., et.al., "Synchronization of the NTT Digital Network," ICC-1976, pp. 25/10-15.
4. "Experimental Packet-Switched Service - Parts 1, 2 and 3," The P.O. Electrical Engineering Journal, 1975.
5. Pouzin, L., "Presentation and Major Design Aspects of the Cyclades Computer Network," Third Data Communications Symposium - 1973, pp. 80-87.
6. Thompson, M., "Switched Networks of the DCS 1980-1982," ICC-76.
7. Kumerle, K., "Multiplexer Performance for Integrated Line and Packet Switched Traffic," ICC-74, Stockholm, Sweden, W1227.
8. Coviello, G. and Vena, P., "Concept for an Integrated Circuit and Packet-Switched Telecommunications System," DCA Technical Comment No. 8-75, Jan. 1975.
9. Coviello, G. and Vena, P., "Integration of Circuit/Packet Switching by a SENET (Slotted Envelope Network) Concept," NTC-1975, pp. 42/12-17.
10. Winkler, G., "Path Delay, its Variations, and Some Implications for the Field Use of Precise Frequency Standards," Proc. IEEE, Vol. 60, No. 5, May 1972, pp. 522-529.
11. Hewlett Packard 1977 Electronic Instruments and Systems Catalog.
12. Allan, D., "Statistics of Atomic Frequency Standards," Proc. IEEE, Vol. 54, No. 2, Feb. 1966, pp. 221-230.
13. Pierce, J., "Synchronizing Digital Networks," BSTJ, March 1969, pp. 615-636.
14. Stover, H., "Communications Network Timing," DCEC Technical Report No. 43-75, Sept. 1975.
15. SENET-DAX Study Final Report, DCA Contract No. DCA100-75-C-0071, GTE Sylvania, Needham, Ma., June 1976.
16. Rosner, R., "Large Scale Network Design Considerations," Proc. ICC-1974, Stockholm, Sweden, Aug. 1974, T2258.
17. Nielsen, P., "Some Optimum and Suboptimum Frame Synchronizers for Binary Data in Gaussian Noise," Trans. on Com., June 1973, pp. 770-772.
18. Mallory, P., "Methods of Frame Sync Analysis," NTC-71, pp. 264-269.
19. Murphy, J., "Some Results on Frame Synchronization," IEEE Summer General Meeting, Toronto, Ontario, June 1963, CP 63-1161.

RESULTS OF INVESTIGATIONS FOR THE CLOCK FREQUENCY CONTROL
AND DISTRIBUTION SYSTEM IN THE DIGITAL TELEPHONY AND DATA
NETWORKS OF THE DEUTSCHE BUNDESPOST AND FUTURE PLANS

W. Ruediger Slabon

Fernmeldetechnisches Zentralamt
der Deutschen Bundespost
Darmstadt
Federal Republic of Germany

Summary

This contribution outlines the state of development in the field of synchronization and timing requirements for the digital data and telephone network of the Deutsche Bundespost.

A comparison of reliability, stability and accuracy of the different synchronization methods is followed by the description of the system which is a present investigated with a view to a possible introduction into the Deutsche Bundespost network.

The digital network operated synchronously according to the single-ended phase averaging method is made accurate by use of the standard frequency system (reference oscillator) of the Fernmeldetechnisches Zentralamt. The tuning method used for this purpose was developed for FDM transmission systems. (Distributed master system).

Some particular questions such as delay variations, accuracy of the control frequency devices etc. have been studied by field-trials as well as the particular problems of the single-ended synchronization system. The results are reported briefly. They show that the envisaged system seems to be the most appropriate and economic solution for the Deutsche Bundespost network.

Introduction

This paper will deal with the synchronization and clocking requirements and the application of different clocking systems to digital networks rather than the clock devices themselves. The brief overview may give especially the designers and manufacturers of oscillators additional information about the specific needs of digital networks.

Accuracy of Digital Networks

To avoid loss of information by bit slips in an integrated digital network or to reduce it to a tolerable value, the clock frequencies of the digital switching centres must operate synchronously or plesiochronously respectively.

According to C.C.I.T.T. Recommendation G.811, the maximum frequency deviation in a national digital network should not exceed ± 1 frame (256 bits) in the 2,048 kbit/s system within 70 days. In a plesiochronous network this value corresponds to an oscillator accuracy of 1×10^{-11} .

Synchronization methods

General

The above mentioned requirement can be achieved by means of one, or a combination, of the three methods known to date, viz.

1. mutual synchronization
2. master-slave synchronization
3. plesiochronous operation mode.

The individual methods are different with regard to reliability, accuracy, stability and complexity of the network arrangement. Among these reasons the reliability of the method plays a decisive part in the selection of the appropriate synchronization method. In the following these three methods are discussed.

Mutual Synchronization

1. Since all network nodes contribute to the synchronization of the entire network, frequency deviations (e.g. aging) of one node affects the vicinity of another node.
2. Temperature-dependent delay variations affect the frequency and phase position of the network nodes without impairment of the network synchronism as long as they are compensated for by a buffer store.
3. The outage of one nodal oscillator does not influence the synchronizability of the network provided that this node contributes no longer to the synchronization.
4. The interconnection of networks may lead to difficulties because the networks or isles of networks may differ in their frequencies although there is synchronous operation among them.
5. The complexity of the system devices is small.

Master-Slave Synchronization

There is assumed to be star-shaped connections between all nodes of the network and a master node or a master clock.

1. If clock equipment as Caesium and Rubidium atomic standards is provided in the master node, there is no considerable frequency deviation in the network. Hence, the integration of additional networks does not cause any problems.
2. Failure of a slave node does not affect the rest of the network.

3. Failure of the master node gives rise to high slip rates in the network
4. Failure of a clock distribution link causes frequency inaccuracy in the disconnected parts of the network.
5. If for safety reasons two clock distribution systems are set up on different routes the costs of the clock system rise considerably.

Plesiochronous Operation Mode

In these networks all nodes are equipped with high-precision atomic standards. The deviations among the various networks are within permissible tolerances.

1. High reliability of the network is achieved because a failure of a node does not affect the rest of the network.
2. At present the system is very expensive, not only because of the higher investment compared with other methods but also for the larger amount of supervision and maintenance for the many atomic standards in the network.

Synchronization Method for the Digital Networks of the Deutsche Bundespost

Basic

Balancing this evaluation we found that none of the above mentioned systems may fulfill the requirements of the digital network. Therefore the Deutsche Bundespost investigates at present a combined method (Fig. 1). It consists of a star-shaped distribution of a control frequency from a central reference clock generator, which serves as a master frequency with a higher weight to achieve the required frequency accuracy in the digital networks operating according to the single-ended phase averaging method. This method offers the advantages of the master-slave network and of the mutual synchronized network. The disadvantages of the two methods are eliminated to a large extent.

The standard frequency system (stage I)

The standard frequency system of the FTZ 1,2 provides the standard frequency for various services of the Deutsche Bundespost. Daily measurements to compare with the emission of the transmitter DCF 77 at Mainflingen, which emits the frequency and time standards of the Physikalisch-Technische Bundesanstalt (PTB) at Braunschweig, insure that the standard frequency provided by the FTZ has an accuracy of $\pm 5 \times 10^{-12}$. Thus it meets the international requirement laid down for a reference clock in C.C.I.T.T. Rec. G.811.

Standard Frequency distribution (stage II)

With the introduction of the multi-channel 60-MHz FDM system use will be made of an improved control frequency device for carrier frequency synchronization 3,4. The reference frequency (300 kHz), which originates in the FTZ, will be transmitted to all network

levels. Under normal operating conditions it is ensured that, at the adapting unit of the digital switching centre, the reference frequency provided via by-pass circuits is directly available to the digital network as a weighted master frequency. In case of fault, the control frequency device causes the change-over to an oscillator which is controlled by a precision crystal. Use is made of a store so that in the absence of the reference frequency the last control voltage applied is not lost. Thus, the relative frequency deviation remains $< 1 \times 10^{-9}$.

Mutual synchronization of the digital switching centres (stage III)

In this system the line clock is extracted from a certain number of selected incoming PCM links and its phase is compared with the switching centre clock. The length-modulated pulses of the flip-flops measuring the phase difference are added via a resistance network and smoothed via a low-pass filter with a specific time constant. By the resulting output voltage the frequency of the nodal oscillator is continuously tuned so that it approaches the average value of the incoming frequencies (Fig. 2). In the transient state the frequencies of all nodal oscillators are equal 5,6,7. In order to prevent the network frequency from drifting by a change of frequency at individual nodes, all synchronous network isles are synchronized by the master clock with the higher weight.

If the master clock fails, synchronization among the network nodes is kept up by phase averaging. The network drift experienced in this case depends on the duration of the master clock failure and the accuracy of the crystal clocks. This arrangement does not require duplication of the clock distribution system on the link between master oscillator and the digital parts of the network for an increase of reliability.

Results of Investigations

Mutual Synchronization

1. Experimental arrangement

A field trial has been carried out since the middle of 1972 to study in practice the phase averaging method according to the single-ended principle. For the experiment, PCM transmission lines between the local exchanges at Pasing and Puchheim in the Munich area have been intermeshed as shown in Fig. 3. The experimental arrangement consists of the network with the clocks 1...4, which synchronize each other, and the equipment for measurements and registration.

Network data

- cable length: approximately 12 km
- line delay time and line delay time variation: $60 \mu s \pm 300 \text{ ns}$
- number of repeaters per PCM transmission line: 4

Oscillator and control arrangement data

- accuracy of free running oscillator: $f/f = \pm 5 \cdot 10^{-8}$ at 2.048 MHz

- aging: initial value $3 \cdot 10^{-8}$ /month (preaged 1 month)
- tuning range: $\Delta f/f = \pm 2,5 \cdot 10^{-7}$
- working range of the phase comparator: ± 16 bits at 2.048 MHz (Frequency dividers enlarge the working range of the phase discriminators)
- loop gain: 0.05 Hz/bit at 2.048 MHz

2. Results

Establishment of the network

- First intermeshing of the network :
It is not necessary to observe certain build-up rules such stepwise build-up of the network etc.

- Enlargement of the network is no problem.

Normal operation

- Required working range of phase comparator could be proved to be sufficient, the working range of the phase comparators was enlarged by frequency dividers, i.e. binary counters. When the working range is exceeded, these counters are reset in such a manner that the phase discriminators are reset into the mid-position of their working range. In the following, this process is called reference phase switching.
- Line delay time variations occurring during a year have had only little influence on the phase relations in the network. The working range of the phase comparators was not exceeded.

Disturbed operation was brought about intentionally by manual variation of the frequency of one oscillator, phase reference switching by hand, and disconnection of clock inputs which corresponds to an interruption of PCM transmission lines.

While simulating the aging process of a quartz oscillator, three different stages were to be distinguished:

- The frequency of the oscillator was inside the control range of the others: the network remained stable.
- The control range was just exceeded (deviation from nominal frequency: 0,8 Hz): Reference phase switching occurred approximately every 40 s. The phases of the remaining network showed a jitter of up to about 14 bits (lines starting and ending at the disturbed oscillator) and up to 4 bits (other lines). Hence, lack of synchronism is confined to lines originating at the disturbed oscillator and to lines arriving from the undisturbed part of the network and connected to the disturbed node (Fig. 4).
- The control range was considerably exceeded. When the control range is exceeded even more, there will be less phase shifts in the rest of the network because of the low bandwidth of the control circuits. Also, the frequency jitter of a node diminishes as a function of the number of nodes separating it from the disturbance

(lower Fig. 4).

Reference phase switching was induced intentionally. It influences the phase of PCM information (absolute phase). Fig. 5 shows the influence of reference phase switching processes which are statistically distributed over the whole network. Fig. 6 shows the influence of repeated phase switching at one input. After one or two reference phase switching processes the phase discriminator does not change its position any longer. The change of absolute phase remains very small (maximum of 3 bits). These test showed that the system satisfies the requirements for the synchronization of regional areas and provides even in the case of some disturbances still acceptable service.

Standard Frequency Distribution

The frequency deviation corresponds to the requirements of the FDM systems for which the control frequency device was developed. First considerations of the achievable accuracy have, however, shown that an accuracy of $1 \cdot 10^{11}$ will be achieved in the digital network.

A field trial being performed in the Deutsche Bundespost network will, before long, permit a definite statement to be made on the usefulness of this method for the provision of the digital network.

The experimental arrangement provides the transmission of the reference frequency (300 kHz) by FDM-transmitting systems over distances up to 1000km using alternatively radio links or cables. Results preliminarily show the following behaviour:

- Long term accuracy measured: $\approx 1 \cdot 10^{+12}$
- Phase jumps : none
- Phase deviations : ≤ 40 ns (periodically)
- Security (measured over 2 month periods):
Short-term interruptions (100ms): 30 ... 40
Long-term interruptions (1s) : 20 ... 30
- Accuracy in case of interrupted master:
Within 24 hours: $\approx 1 \cdot 10^{-10}$

This would in a data network, where the aligner at a 64 kbit/s input has a capacity of e.g. 8 bits, result in a slip after a period of about 8 days.

Although the field trial is still continued these first results satisfy completely the requirements of digital networks. The short- and long-term interruptions will not affect the digital network, since it relies only for the long-term accuracy on the standard frequency distribution system.

Implementation of the Synchronization method

Setting up of digital switching centres in the Deutsche Bundespost network is intended to start in the eighties. To begin with, small isles comprising 2 to 4 digital

switching centres will be established, which will already be synchronized according to the above-mentioned method. At present, the reference frequency network provided for the FDM systems covers already 80% of the nodal exchanges. Therefore, it will always be possible to apply the master frequency to the nodes of the first network isles down to the nodal exchanges.

Later on, these network isles will be interconnected to form digital network areas. At the lower level of the communication network digital transmission will prevail and will be integrated into the digital switching systems.

At the upper level of the communication network the FDM systems will gradually be replaced by multi-channel digital transmission systems. Instead of being tuned via FDM channels the digital network would then be synchronized direct by the NFA/FTZ or by several local NFAs distributed in the network.

Problems still to be investigated

The most important unsolved problems are:

- A permissible value ($\Delta\varphi_{\max}$) for the magnitude of the phase shift must be specified. The same problems will also be encountered on digital long distance transmission systems.
An interruption of the master clock causes the control frequency device to initiate a change-over to the crystal. The frequency deviation f_q of the crystal has to be determined by means of tests performed in the network.
- The problem arising in conjunction with the interconnection of digital network isles (different phase positions $\Delta\varphi_1 + \Delta\varphi_n$) shall be investigated during a field trial to be carried out on the links of the PCM data network (beginning about 1978).

In spite of some unsolved problems the method described seems to be suitable for national telephone and data networks. On account of its high accuracy of $> 1 \times 10^{11}$ and the high immunity against failure it will be possible to integrate the digital network of the Deutsche Bundespost without difficulty into international telephone and data networks.

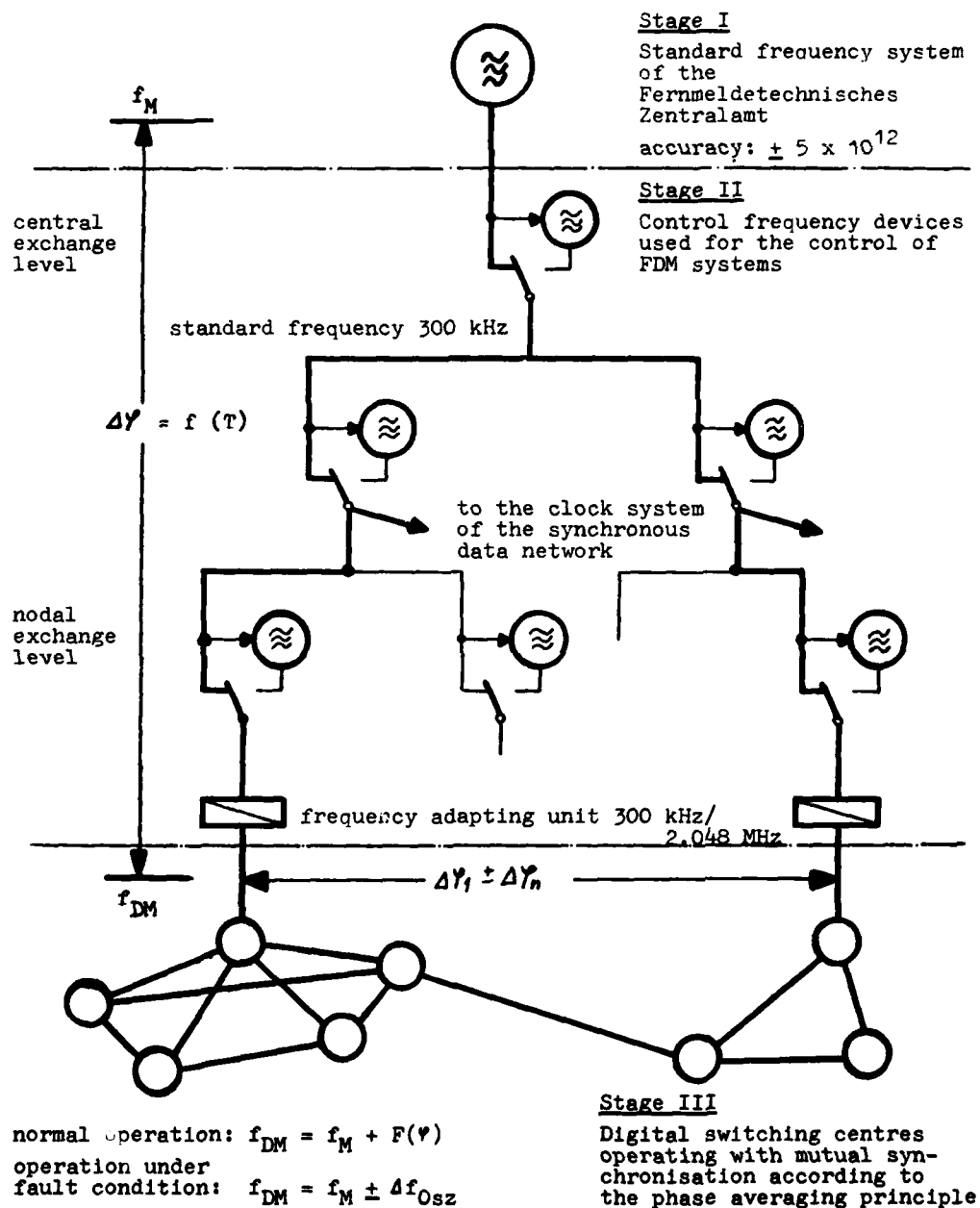
References

1. Übertragungswege der Normalfrequenz der Deutschen Bundespost
J. Bastelberger
Der Fernmeldeingenieur Vol. 27 No. 11, 1973
2. Normalfrequenzen, ihre Übertragung und Messung
J. Bastelberger
Der Fernmelde-Ingenieur Vol. 22 Nos. 5 und 6, 1968
3. Steuerrfrequenzerzeugung für TF-Systeme bis 60 Hz
H. Löffler
Elektrisches Nachrichtenwesen, Vol. 50 No. 1
4. Die Frequenzgenauigkeit in der Trägerfrequenztechnik

G. Pankow

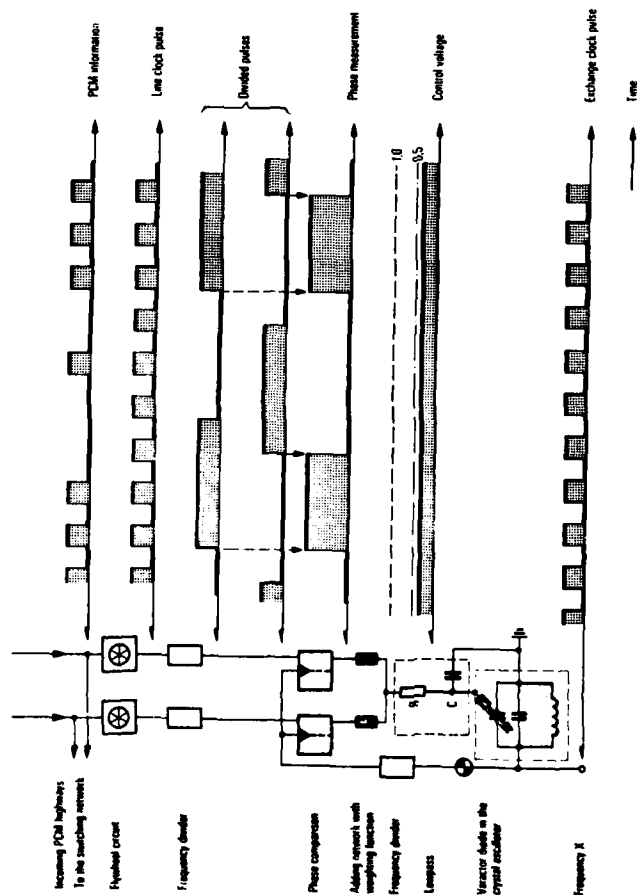
Taschenbuch der Fernmeldepraxis (1974)

5. Synchronisationsverfahren nach dem Phasemittelungsprinzip
O. Karl
NTZ 28 (1975) No. 12
6. Gegenseitige Taktsynchronisation in integrierten PCM-Vermittlungsnetzen
A. Darré und O. Karl
Information Fernsprechvermittlungstechnik 7 (1971) No. 1
7. The Single-ended Principle of Digitally Controlled Phase Averaging for the Synchronization of Communication Networks
H. L. Hartmann
NTZ 28 (1975) No. 12

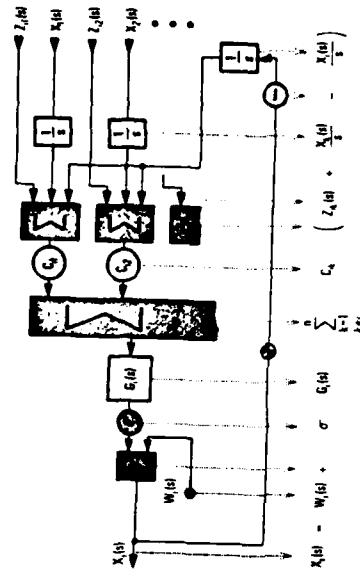


Principle of clock distribution from the central master clock to the digital network level by means of the distributed master method

Fig. 1 Synchronization of the digital networks of the Deutsche Bundespost

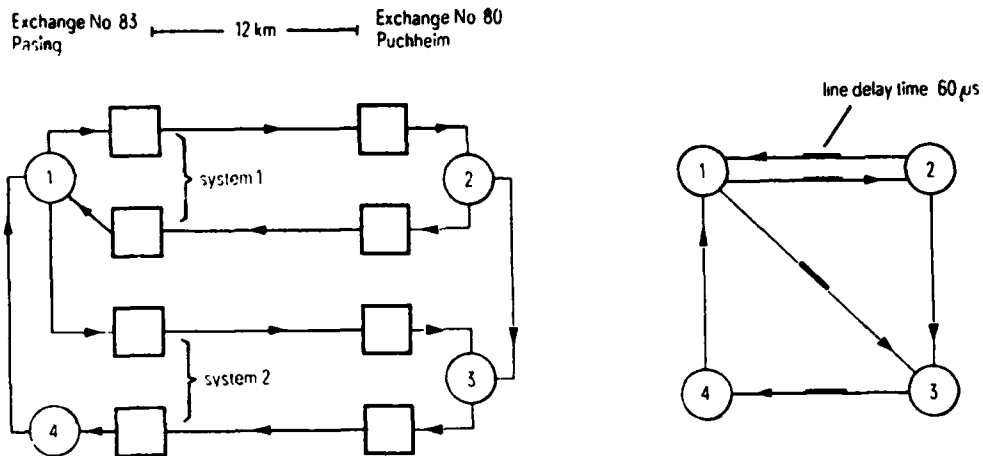


2a. General diagram of a network node. The control voltage is obtained through a phase measurement. The other nodes only supply the PCM information, but no information on the control condition there, designates the point at which the output variable is fed back for comparison with the input variables



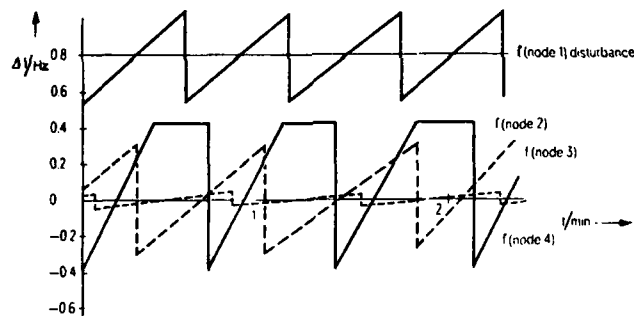
2b Mathematical model of the circuit shown in Fig. 2a after application of the Laplace transform. This can be used to derive equation (1) for the i node.

Fig. 2 Mutual synchronization system using the single-ended method

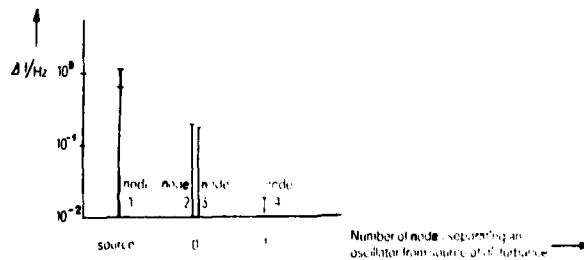


Configuration of experimental network

Fig. 3 Synchronization field trial



a) Frequencies of the 4 nodes when the frequency of oscillator 1 is out of the pull-in range of the remaining oscillators $\Delta f = 0.8$ Hz



b) Frequencies of the 4 nodes versus distance from source of disturbance $\Delta f = 0.8$ Hz

Behaviour of the network when one oscillator is disturbed

Fig. 4 Influence of disturbances on the synchronized network

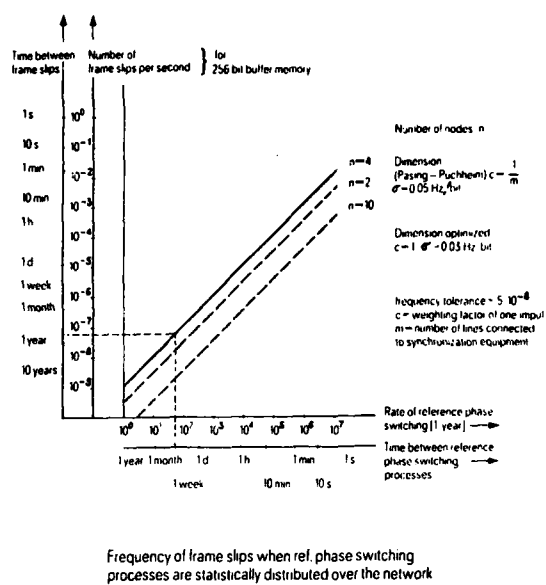


Fig. 5 Frame slips in digital networks

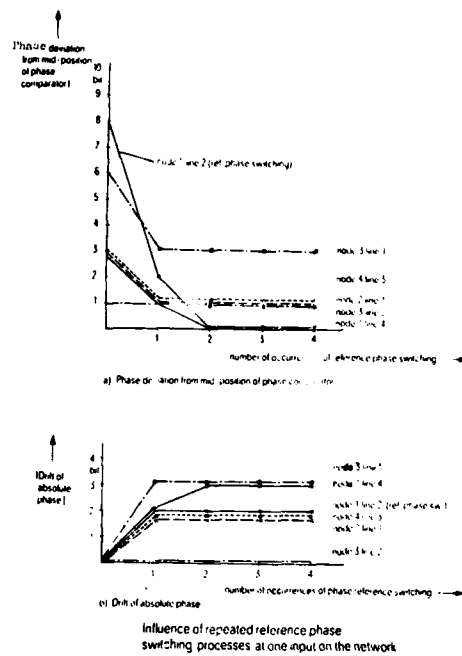


Fig. 6 Reference phase switching

SYNCHRONIZATION OF A DIGITAL NETWORK USING
A POSITION LOCK AND HMS SYSTEM

P.K. Verma
Bell Canada, Ottawa

C. Lemieux
Bell Canada, Ottawa

and

H. Daams
National Research Council, Ottawa

Abstract

This paper briefly describes the synchronization system for the Dataroute, a nationwide digital data transmission network of the Trans-Canada Telephone System, operating since April 1973. Recent operating experience and measurements on the network as regards to the accuracy and stability of the master clock frequency are also discussed.

Introduction

The Dataroute,¹ a Trans-Canada digital data network, was established in April 1973 to meet the transmission requirements of the growing computer communications market from coast-to-coast in Canada. The initial synchronization system of The Dataroute was based on a Master-Slave system using a station synchronizer with two fixed priorities for determining the appropriate source for running the local nodal clock. Plans were made simultaneously to elevate the network synchronization to one with a Hierarchical Master-Slave capability within a short time period.^{2,3,4}

The decision to utilize the HMS feature was based on considerations involving:

- a) ease with which network growth, both addition of new nodes and new circuits to existing nodes, could be accomplished with an adaptive, rather than a fixed priority scheme;
 - b) enhancement of network security;
- and
- c) relative insensitivity to link failure of both short and long term duration.

The Dataroute System

The Dataroute Network is a transmission network linking some thirty-five cities by 56 KBPS or 1.544 MBPS digital transmission facilities.

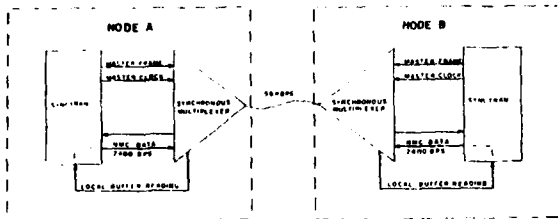
The Dataroute is a Master Frame system. (A Master Frame system is one where the demultiplexing Frame is common to all multiplexers in a node and is provided by a common equipment).

Selection of the 'best' stream at any node from the point of view of synchronization is effected by associating a travelling signature along with each bit stream. Having determined the best stream, however, the local clock does not frequency lock to it. Instead, a positional information with the remote multiplexer is exchanged in order to derive an error signal to control the local clock.

Between any pair of communicating multiplexers, provision is made for exchanging the receive buffer readings. The difference between the buffer positions generates an error signal which is used to effect a movement of the nodal frame and the nodal clock in a direction to nullify the difference. This is shown in figure 1.

It can be easily seen that a constant difference between the two (local and remote multiplexer) buffer readings would amount to the nodal clock rate being identical to the clock driving the remote multiplexer. In addition, if the buffer readings are identical, a real time alignment of the two nodal frames has also been achieved. This alignment process can continue to take place between pairs of nodes until all nodes are similarly aligned. Hence, the process of maintaining real time alignment of the different nodal frames effectively achieves synchronization. The attainment of synchronization is thus based on exchange and control of buffer position information qualifying the synchronization system as a position control system.

The network frequency is continually measured for accuracy and stability against standard clocks maintained at the Time and Frequency Laboratories of the National Research Council of Canada in Ottawa. The measured frequencies have been found to be within the acceptable range of operation.



TWO NODE CONFIGURATION DATAROUTE SYNCHRONIZATION
FIGURE 1

Hierarchies of Multiplexing System

Hierarchies of multiplexing systems provide an interesting situation for the provisioning of position lock clock derivation schemes. In such systems, there are two possible options.

- a) The positional information may be exchanged amongst the highest order multiplexers. In this case, the lower order multiplexers have no receive buffer and the derivation of clock frequency is done entirely by the positional information exchange between the higher order multiplexers.
- b) The synchronization information exchange may be based on the lower level multiplexer. In this case two options are available:
 - i) the synchronization equipment receives the buffer readings of both the higher and lower level multiplexers. Equalization is first made on the lower level multiplexer which is usually 'coarser'. Subsequently, the finer equalization can be made based on the buffer readings of the higher level multiplexer.
 - ii) the higher level multiplexer can be made non-master frame.

In the Dataroute system, the signature imbedded in the bit stream is demultiplexed by the 56 KBPS multiplexers. The higher level multiplexer is made non-master frame thus eliminating the need for communicating buffer readings of the 1.544 MBPS multiplexer.

Summary

This paper has presented details of the synchronization system of the Dataroute, a Trans-Canada digital data network operating since April 1973. The synchronization system is a Hierarchical Master Slave System with provision to align in real time the demultiplexing frames at all the nodes.

References

1. An Overview of Dataroute: System and Performance, D.J. Horton and P.G. Bowie, presented at the International Conference on Communications, Minneapolis, June 1974 PP. 2A-1 to 2A-5.
2. J.G. Baart, S. Harting and P.K. Verma, Network Synchronization and Alarm Remoting in The Dataroute, IEEE Trans. Commun., Vol. Com-22, PP 1873-1877, Nov. 1974
3. S. Harting and P.K. Verma, "Universal Time Frame: A New Feature for Delay Minimization," IEEE Trans. on Commun., Vol Com-23, PP 1339-1342.
4. P.K. Verma and A.J. Delorenzi, "Operating Experience with the Canadian Dataroute Synchronization System, presented at the International Conference on Communications, Philadelphia, June 1977, PP 25-16 to 25-20.

AN OVERVIEW OF TDMA FOR DIGITAL SATELLITE COMMUNICATIONS AND THE CENSAR SYNCHRONIZATION EXPERIMENT

K.E. Brown and P.P. Nuspi
Department of Communications
Communications Research Centre
Ottawa, Canada

Summary

The concept of Time-Division Multiple-Access satellite communication is described indicating the need for highly accurate time synchronization of all participating stations. Factors affecting the propagation time and contributing to the achievable accuracy of synchronization are discussed. The design and implementation of a unique synchronization system called CENSAR are described. Synchronization to within ± 10 ns has been recently demonstrated on the Hermes (CTS) satellite using this method.

Some future trends in satellite communications are summarized.

Introduction

In modern satellite communication systems, it is common practice to place the satellite in a near-circular orbit around the earth above the equator at an altitude such that the orbit period is 24 hours. Such a satellite is termed geo-stationary since, to an observer on the ground, the satellite appears virtually stationary. This simplifies antenna tracking and increases the satellite availability at any point within the coverage region to nearly 100%.

To access the satellite, an earth station processes the baseband signal from the terrestrial interface, modulates the result onto a carrier and up-converts to a suitable radio frequency (Figure 1). This modulated signal is transmitted to the satellite where it is amplified, shifted down in frequency and re-transmitted to the earth. The receiving earth station amplifies, down-converts and recovers a signal which is an approximation to the original baseband signal (since the link is not perfect and is subject to noise and interference the received signal is not an exact replica of the transmitted signal).

Several frequency bands have been allocated internationally for satellite communication based on such factors as penetration through the ionosphere, spectral absorption, interference, frequency sharing, etc., as well as non-technical reasons. Three common pairs of frequency bands are:

1. 6 GHz for the up-link and 4 GHz for the down-link - these frequencies are used in the majority of commercial communication satellites today - these frequency bands are shared with terrestrial microwave relay;
2. 14 and 12 GHz respectively - these frequencies are beginning to be introduced in experimental satellites and are planned for use in new generation satellites, 12 GHz is also allocated to the broadcasting satellite service; and
3. 30 and 18 GHz - for use in future satellites.

The baseband signal can be anything from a single 4-kHz-wide voice channel through various analogue or digital multiplexed signals to 6-MHz-wide television and even to the higher levels of multiplexing up to 274 MHz in the future.

Other frequencies in the UHF and SHF bands are used for broadcast, navigation and government purposes.

System Impairments

The radio path from the earth station to the spacecraft traverses the troposphere and the ionosphere. These media introduce certain modifying effects which must be taken into account. For example, in the atmosphere the velocity of propagation and the path length of the radio ray are functions of the temperature, pressure, humidity and turbulence profiles, and the signal level is a function of the water, snow and ice content and, at certain frequencies, of the gaseous content. In the ionosphere both the velocity of propagation and the signal level are dependent upon the electron content.

A geo-stationary satellite is not normally perfectly stationary with respect to the earth due to, for example, station drift and the gradual change of orbit inclination and eccentricity caused by gravitational and radiation pressure perturbations. These effects generate a drift within a small volume of space whose largest dimension is typically about 140 km with a maximum relative velocity of the order of 1 m/s. This motion gives rise to a doppler shift in both up- and down-links. In the time domain, the drift motion causes a diurnal variation in the transmission time from an earth station to the satellite of up to 250 μ s. Commercial satellites are usually station-kept to these tight bounds. The measured diurnal variation for Hermes, which has no North-South station keeping was 650 μ s in the spring of 1976.

Multiple Access

In utilizing a satellite for communications, it is generally necessary from both economic and operational viewpoints to plan to share the expensive resources in the spacecraft with many earth station facilities. There are several ways in which this can be accomplished. This paper is concerned with time sharing in which the full power and bandwidth of the satellite transponder are allocated sequentially on a cyclic basis to each participating earth station for a short period of time.

Voice communication is the major constituent of all present day communications. For compatibility with digitized voice, each earth station generally has access to the satellite once each 125 μ s or multiples thereof (125 μ s corresponds to the minimum Nyquist sample rate for a 4 kHz wide voice channel, a bandwidth which is adequate for telephony). This process, termed Time-Division Multiple Access (TDMA), requires accurate synchronization. By way of an example, the CENSAR synchronization scheme currently under investigation at the Communications Research Centre in Canada will be described. Earth station burst transmissions must arrive at the satellite without overlap and without large idle intervals between the bursts. This time synchronization must be maintained under the influence of satellite drift and the ionospheric and atmospheric effects mentioned above. Also, frequency synchronization must be maintained under the influence of doppler shift and frequency drift in the up and down conversion chains and in the satellite frequency translation.

TDMA

The prime reason for considering time-division multiple access is that this method makes most efficient and effective use of the satellite capacity. In addition it is cost-competitive with other multiple access techniques. Other advantages include adaptability to changes in demand as well as all the many reasons for the use of digital signals, e.g. rugged signals capable of regeneration and remodulation, use of coding for increased fidelity, compatibility with emerging terrestrial digital facilities, and the feasibility of more complex satellites with on-board processing. TDMA has been under investigation for over a decade¹ by many researchers throughout the world (see table in Figure 2). However at the present time only one system is operational, namely that of Telesat and this system with only two accesses does not demonstrate the full capability and power of TDMA².

Problems involved in the introduction of TDMA include the slow introduction of terrestrial digital transmission and digital switching due to the vast investment in analogue plant, a lack of international standards, and a slowdown in the rate of increase in demand for telecommunications facilities. The TDMA concept however offers sufficient potential benefit that research in this area continues at a high level. Recently Intelsat issued a request for proposals for a TDMA test bed^{3,4} and the European countries are planning extensive pre-operational TDMA experiments on OTS in 1979.

TDMA General Concept

As mentioned previously, TDMA refers to the time-shared use of a satellite transponder by means of short bursts of information from each earth station directed to one or more receiving stations. The system must be synchronized such that (a) the bursts do not overlap or have large idle periods between them and (b) proper burst identification must be possible.

To implement a TDMA synchronization system, two conditions must, in general, be fulfilled:

1. All participating terminals must operate on a common time-base;
2. Each participating terminal must know its exact transmission delay to the satellite.

The first condition is fulfilled by transmitting a timing signal from a control terminal to all participating terminals. In the CENSAR system, the second condition is achieved by, in effect, determining the satellite position at the control terminal and distributing the positional information via the satellite to the participating terminals. Each terminal then calculates its delay to the satellite. As the satellite moves, the positional information is continuously up-dated so that each participating terminal can compensate.

Shown in Figure 3 are the essential features of a TDMA timing hierarchy. A frame is the time interval selected for an access assignment plan to repeat, a superframe is a group of frames in which is distributed internal signalling information. Each frame in general consists of one or more reference bursts followed by a string of message bursts; due to imperfect synchronization a guard time is usually required between bursts. A burst is a modulated carrier signal which occupies an assigned time-slot within a frame. A reference burst or a sync burst occurs at the beginning of the frame and often has special structure. A message burst typically consists of a preamble, the

message portion and, in some systems, a "postamble". The preamble may consist of a portion for carrier recovery, a portion for symbol timing (clock, bit-timing) recovery, a burst codeword for burst synchronization, a station-identification code, and other house-keeping portions. The message (information) portion contains traffic which may be multiplexed and sometimes coded. In some applications, a postamble of a few symbols is used for decoder or demodulator quenching. A symbol is one of the possible modulated carrier signals for the burst and occupies a symbol length defined by the clock rate in the burst. The transmission rate is the instantaneous bit rate of the burst. The number of accesses can in theory be anywhere from two (as in the case of Telesat) to several hundred.

In order to reduce the system overhead it is obviously necessary to minimize the duration of the allowed guard time and of the pre- and post-amble. Typical TDMA systems have guard times ranging from less than 20 ns to over 200 ns. Time-slot synchronization must thus be dynamic to accommodate satellite drift, atmospheric and ionospheric effects, and equipment temperature dependence. Synchronization must also be very accurate for maximum efficiency and must be maintained under burst conditions. Clock and symbol-timing recovery and burst identification must be accomplished very quickly at the beginning of each burst in order to keep this portion of the overhead to a minimum and also in order to extract the appropriate message portion with minimum error.

Additional equipment is required to compress the continuous signal presented at the terrestrial interface into burst format and to expand the received bursts into continuous format for return to the terrestrial network. The entire process must be accomplished without significant loss of fidelity in order to preserve the required quality of service. The TDMA process thus calls for very sophisticated circuitry.

Concept of Centralized Synchronization

The centralized synchronization concept evolved from an original idea of de Buda⁵. Consider one earth station which transmits a short ranging burst to the satellite and receives the same burst after the appropriate two-way transmission delay - about a quarter of a second. The station measures the time interval between transmission and reception and the satellite must thus lie on a sphere centred on the earth station with a radius (in time units) equal to half the round trip delay. Thus two other identical stations, suitably distant, define two more spheres and the satellite then lies at the intersection of these three spheres. The required synchronization accuracy for guard times of the order of 20 ns must be better than one part in 10^{10} . By coding the ranging bursts the satellite position can be broadcast to all other stations in the network.

The above process presupposes that each station can receive its own transmissions. For spot-beam satellite this is not generally the case and it is thus necessary to modify the approach. Consider now a pair of earth stations in different spot beams. Station A transmits a burst to the satellite which is routed to station B, station B acknowledges by sending a burst back through the satellite to station A. Station A measures the time duration between transmission of its burst to reception of B's burst. Station A can now predict that the satellite lies on the surface of an ellipsoid of rotation with A and B as focii. Ranging to two other stations of type B forms two other ellipsoids each having A as one focus. The satellite is again accurately located as the

intersection of three ellipsoids. Station A now has all the required information and, by suitably coding its ranging bursts, can broadcast this information to all other stations in the network.

The CENSAR System

The CENSAR (CENTralized Synchronization And Ranging) system is a new method of time-division multiple-access synchronization for spot beam application using the concept outlined above.

The system offers centralized control and ranging data processing. The synchronization equipment at the remote stations is thus quite simple and hence there is a potential cost benefit over other TDMA synchronization systems. In addition the CENSAR system is highly efficient requiring guard times less than 20 ns and small burst preamble overhead. The potential frame efficiency (defined as the ratio of time allocated to message information to the total frame time), assuming no coding or postamble requirements and traffic quality satellite channels, has been computed at 97% for 30 accesses using the minimum possible frame length of 125 μ s. Higher efficiency would result for a longer frame; e.g. for a 750 μ s frame and 30 accesses the efficiency would be 99.5%. The cost is in increased buffer memory.

An experiment with the Hermes satellite was designed to test the CENSAR concept. Equipment was designed and constructed by Canadian Marconi Company to CRC specifications for this purpose.

The system specifications were set to be compatible with the constraints imposed by Hermes and by the small portable earth station antennae which were designed primarily for TV reception and have limited communication capability in the transmit mode. These stations are used in a variety of experiments, both technical and social.

Five terminals were specified and constructed; these consist of:

1. a central control terminal;
2. three ranging terminals; and
3. an associate terminal with both ranging and simulated message capabilities.

The required software is implemented on a PDP 11-40 computer at the central control terminal and an auxiliary PDP 11-10 computer at the associate terminal; the latter may subsequently be replaced by a microprocessor.

The software provides all the processing and real-time computations through a set of algorithms⁶ using double precision. The system can be reconfigured from the central control and a simple teletype order-wire is incorporated into the system.

An offshoot of the synchronization experiment is an orbit perturbation study which capitalizes on the high resolution of the range measurement process to produce satellite orbit data for subsequent off-line analysis.

Description of the Principles of the Experimental System

The central control terminal and the three ranging terminals operate together to measure the position of the satellite. The fifth or associate terminal represents one of the many user terminals in an operational

system and constitutes a system check. A diagram of the terminal configuration is shown in Figure 4.

Hermes has two steerable spot-beam antennae and two transponder channels. During the TDMA experiment, one spot-beam illuminates the control terminal and one spot-beam illuminates all other terminals.

A common time-base is transmitted from the communications control terminal via the satellite to all terminals in the system (Figure 5). At all other terminals a receiver is phase-locked to this time-base to an accuracy of 2 ns + 0.2 ns. The signals from all other terminals pass through a second channel to the control terminal. These latter signals are allotted specific time slots so that there is no mutual interference (Figure 6).

The channel used to transmit the time-base is called the control channel. By coding the time-base signal the control channel also transmits the location of the satellite and other control information.

The system time-base is periodic with period of 640 milliseconds. This value was chosen to be greater than the two-hop delay (which is approximately 520 ms) to permit range measurement and allow sufficient time for computation. Each period is called a cycle (equivalent to a super-frame). It consists of 5120 frames each with a duration of 125 microseconds. The first 15.2-ns pulse in the first frame of the cycle is called start-of-cycle. When the time-base is received at a ranging terminal 'i' ($i = 2, 3, 4$), the start-of-cycle will be delayed exactly $T_1 + T_i$ seconds from the start-of-cycle in the control terminal, where T_1 is the propagation delay from the control terminal to the satellite and T_i is the delay from ranging terminal 'i' to the satellite.

The control terminal transmits estimated values of T_1+T_1 , T_1+T_2 , T_1+T_3 , T_1+T_4 . Ranging terminal 'i' receives this signal, regenerates the system time-base using the received start of cycle and transmits a time-mark pulse burst at $2(T_1+T_i)-K_i$ seconds in advance of the next received time-base start-of-cycle signal (where the K_i are fixed constant delays). If the control terminal has estimated T_1+T_i correctly, the time-mark from the ranging terminal will arrive at the control terminal K_i seconds after the start-of-cycle. Suppose the initial estimate of T_1+T_i by the control terminal is too large by an amount Δ , then the time-mark will arrive at the control terminal $K_i-2\Delta$ seconds after the start-of-cycle. This is measured by the control terminal. The time estimate T_1+T_i is then adjusted until the time-mark received from terminal 'i' falls into a pre-assigned time slot.

The value of the propagation times are not expected to change by more than +0.25% because of the near-geostationary nature of the satellite. Therefore, T_1+T_i can always be expressed as a large fixed portion and a small variable portion. Only the variable portion needs to be transmitted.

The control terminal, having measured the three two-hop delays and knowing the location of the ranging terminals, can calculate the position of the satellite and thus its own up-link delay⁶. (This quantity cannot be measured directly since, for spot-beam application, the control terminal cannot, in general, receive its own transmission.)

The propagation delay of a signal transmitted from the associated terminal ($i=5$) to the control terminal via the satellite is calculated as a weighted sum of the delay times:

AD-A088 221

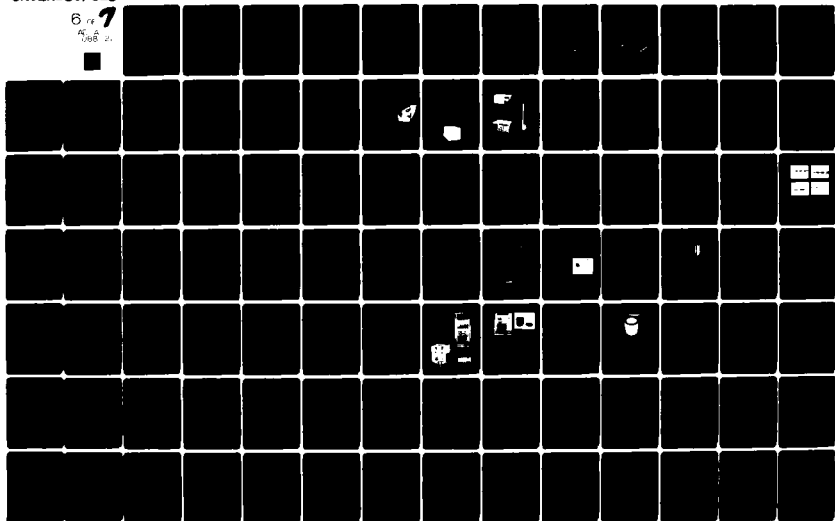
ARMY ELECTRONICS COMMAND FORT MONMOUTH NJ
PROCEEDINGS OF THE ANNUAL FREQUENCY CONTROL SYMPOSIUM (31ST) HE--ETC(U)
1977

F/G 9/5

UNCLASSIFIED

NL

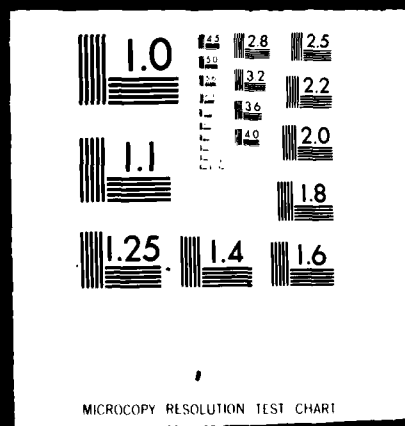
8 of 7
PAGE 11



1 F 120

6 OF 7

AD A
088221



$$D = T_1 + T_5 = d_1 T_1 + d_2 T_2 + d_3 T_3 + d_4 T_4 + d_5$$

where the d 's are fixed constants related to the geographical location of the associated terminal, ranging terminals and control terminal. The associated terminal must advance its transmission by $2(T_1 + T_5) - K_5$ seconds relative to its received time-base where K_5 is a specified constant delay.

Hence, theoretically, a burst of signal transmitted from the associated terminal will arrive at the control terminal precisely within an assigned time slot. A number of associated terminals can be accommodated without changing the basic nature of the system.

System Time-Base Hierarchy

The master time-base in the control terminal is derived from a crystal standard operating at a frequency of 65.536 MHz. Pulses are generated from this time-base to clock the control channel bursts which are transmitted via the satellite to all stations.

The time-base hierarchy is listed in Table 1.

TABLE 1

Time-Base Hierarchy

	Duration	Frequency
1 transmission bit	15.26 ns	65.536 MHz
1 channel = 8 bits	122.09 ns	4.096 MHz
1 frame = 1024 ch.	125 μ s	8 kHz
1 control-bit = 16 frames	2 ms	500 kHz
1 cycle = 320 control-bits	640 ms	1.5625 Hz

The Control Channel

The control channel is a binary channel that carries control and synchronization information from the control terminal to all terminals in the system. This channel has the following functions:

1. distribution of system timing;
2. distribution of satellite position information;
3. distribution of control information.

The control channel occupies the first 31 bits of each frame (473 nanoseconds every 125 microseconds). In this time-slot, one of two 31-bit codes is transmitted. These codes are called mark or space respectively. These codes have been specifically selected such that bit transitions in the code can be extracted by the ranging and associated terminals to give symbol-timing. The time between successive code bursts is exactly 125 microseconds so that frame-timing can also be extracted.

By suitable coding techniques a binary message can thus be transmitted. This message contains information on the variable portions of the four propagation delays, the four relative velocities (used for linear prediction), system status, as well as providing control and order wire.

The high level of coding redundancy ensures an extremely low error rate in the detected information⁷. The equivalent information rate is only 350 bits/s. representing a net redundancy factor of over 700.

System Description - Control Terminal

Figure 7 shows the relationship between the various subsystems in each terminal and the inter-connection between the terminals.

The control terminal has several important functions:

1. It generates and distributes the time-base for the entire system;
2. It measures the two-way delays from each of the three ranging terminals to the control terminal;
3. It calculates the satellite position. (These calculations are performed by a computer located in the control terminal.)
4. It transmits positional information and control commands through the control channel via the satellite to all terminals of the system.

The three ranging terminals transmit in time sequence, each one transmitting in an assigned quarter of the cycle. Hence, the distance-measuring equipment in the control terminal can be time-shared since the transmission from each ranging terminal can be identified by its time-of-arrival in the cycle. The distance-measuring scheme requires the control terminal to operate in co-ordination with each ranging terminal in a closed loop.

The distance-measuring equipment in the control terminal measures the two-way transmission delay to each of the ranging terminals in three steps. Each step for each terminal occupies a pre-assigned portion of the cycle. The first step, called "fine", measures the delay modulo 15.26 nanoseconds to a resolution of ± 1 nanosecond. The second step, called "bit", measures the delay modulo 125 microseconds with a resolution of ± 1 transmit bit (± 15.26 nanoseconds). The third step is called "frame" and measures the delay to a resolution of 125 microseconds.

Ranging Terminal. The function of the ranging terminal is to act in conjunction with the communications control terminal to measure the two-way delay from the control terminal to the ranging terminal and back.

Associated Terminal. The associated terminal represents one of many communications terminals possible in the system. The four range delays received from the control terminal are entered into a computer in the associated terminal. The computer uses these numbers, together with the fixed multiplying coefficients ' d ' to calculate the propagation delay to the control terminal. This value of the delay is used to initiate a message sequence starting at the specified point in the time-base. In an operating system the function of the associate terminal computer can be performed by a microprocessor.

In the experimental system the associate terminal can also perform the function of a ranging terminal. The difference between the measured and the calculated range provides a system check.

Phase I Experiment

The synchronization experiment consists of 2 phases; the first phase was completed in 1976 and the second phase is currently in progress.

The first phase of the experiment incorporated a central control station in Ottawa, Ontario and a separate facility also in Ottawa for the associate terminal. The three ranging terminals were located in: (1) London, Ontario; (2) Rouyn-Noranda, Quebec; and, (3) Quebec City, Quebec. These latter installations together with the main station in Ottawa were shared with other experiments during the experiment interval. The scheduled time consisted of two or three hours every other day for a period of just over two months during which time the equipment, interfaces and software were checked out and satellite positional data logged for future analysis.

Phase I Results

It was found necessary at the beginning of the experiment to "commission" the system - i.e. to refine some of the approximations of equipment and cable delays and, in some cases, to refine the initial estimates of site locations since it was not always possible to survey the site with the equipment in-situ. After commissioning, the measurement error - equivalent to the guard time required between bursts - was found to be less than 15 ns (Figure 8).

As an example of the propagation error reduction, Figure 9 shows the measured diurnal variation for a 24 hour period and the corresponding CENSAR system error on a different scale. It can be seen that a reduction factor of almost 10^5 has been achieved. Further reduction is possible with additional processing but is not warranted for the present applications.

The prediction error consists of basically two components (1) a diurnal sinusoidal component of peak-to-peak amplitude of ± 4 ns which, with sufficient history, is predictable and could therefore be reduced and (2) a noise component with peak-to-peak amplitude of ± 1.5 ns. Some of the contributing factors to this fine structure were anticipated or have been recently identified and these could also be reduced in future implementation. In addition to these components, the contribution from atmospheric conditions can be predicted with reasonable accuracy from easily measured parameters at each ground station - namely pressure, temperature, humidity and antenna elevation angle. It has been calculated that this contribution would amount to about ± 2 ns over a year. The contribution to range error from rain is small and probably not worth the engineering effort and expense to try to compensate for. The contribution from the electron content of the ionosphere is frequency dependent and, for the conventional 4 and 6 GHz satellites, is quite significant. At 12 and 14 GHz the contribution is less than ± 2 ns about nominal at southern Canadian latitudes depending on time of day, sun spot activity, etc. The electron content is a relatively difficult quantity to measure and also difficult to predict with any confidence, and thus it is unlikely that this effect can be compensated

The CENSAR measurement resolution is less than one nanosecond which corresponds to about 5 parts in 10^{11} . The absolute accuracy of ranging is not known since there is no universally accepted yardstick but it is estimated that the accuracy is at least as good as any other existing ranging scheme. Figure 10 shows a real time measurement of the ranges to two stations, this plot indicates the 1 ns resolution and the fact that, since the satellite is moving in three dimensions, the maximum range will occur at slightly different times for different ground stations. (It was not possible to include all four stations on this plot since the chart recorder only had two pens.)

CENSAR Efficiency

The TDMA system overhead as mentioned previously consists primarily of the guard time between bursts, and the burst pre- and post-amble. From the above measurements and from prediction of contributions of factors such as the built-in implementation errors, the electron content of the ionosphere, etc. a guard time between bursts of 20 ns is generous. Clock, symbol timing and frame-synch recovery is built into the system and thus the burst preamble is relieved of these functions and may be significantly reduced over other forms of TDMA. The ranging bursts take up a very small portion of the frame and have an insignificant effect on the frame efficiency. The overhead for CENSAR is thus very small and this system promises a high efficiency communication capability. The centralized synchronization concept has been proved valid and, since the remote ranging and associate stations are relatively simple, practical implementation should be economic.

Phase II Experiment

The second phase of the experiment has separate spot beams, a larger baseline for increased resolution, and minor modifications and improvements to both hardware and software which resulted from the first phase experience. In addition atmospheric conditions at each location are monitored. It is hoped to correlate the delay variations with these atmospheric conditions. This phase has stations in Ottawa in Eastern Ontario, Thunder Bay in North-Western Ontario, and Winnipeg, Brandon and Thompson all in Manitoba (Figure 4).

Future of Satellite Communication and TDMA

As can be seen from the above and from many other studies^{1,8} time-division multiple access of a communications satellite is a viable technique offering great promise for highly efficient, versatile, flexible, and ultimately very economic communication. Further advances both in satellite technology in general and in TDMA systems in particular are foreseen.

The current generation of satellites operate at 4 and 6 GHz and consist basically of up to 12 frequency-translating amplifiers each with bandwidths of the order of 40 MHz. The antenna coverage is global (in fact only 30% of the earth is visible) and the transmit power is limited by the available prime power which, on spin-stabilized satellites, is derived from solar cells on the cylindrical body of the spacecraft. Such satellites have no on-board processing and are transparent to the type of modulation used. Fairly large earth station antennas are required and these must be located remote from centres of population to avoid interference with terrestrial microwave facilities. This necessitates expensive back-haul facilities.

The next generation of satellites will operate at 12 and 14 GHz with transponder bandwidths up to 500 MHz. Spot beam coverage and dual polarization permit power concentration and frequency re-use. Spacecraft will be three-axis stabilized and derive primary power from extendable solar sails which are not limited by the dimensions of the spacecraft body. The resulting higher available power, beam concentration, and a frequency allocation that does not interfere with terrestrial microwave permit smaller earth stations which can be located, for example, on the roof of telephone company central offices eliminating the need for back-haul facilities. Heretofore the designer of this type

In the planning stage are satellites with on-board dynamic switching which can be synchronized to the TDMA system permitting several bursts from one station to be routed via different down-link spot beams to several geographically displaced ground stations. Equivalent routing in a frequency-division multiple-access system is virtually impossible due to the weight of the required channel filters. Also planned are satellites with the capability of signal regeneration and remodulation geared to specific signal formats so that the overall system can be optimized.

In conjunction with these developments TDMA systems are planned to operate at higher speed with features such as:

1. demand assignment whereby the available capacity is assigned according to the traffic requirements at that instant;
2. digital speech interpolation which utilizes the quiet periods between words and whilst the other party is talking to interleave other voice signals thus doubling the available capacity; and
3. sophisticated coding techniques permitting error correction.

On the horizon are satellites operating at 18 and 30 GHz, satellite constellations with intersatellite links - probably via laser, the TDMA synchronization function taken over by the satellite, and the capability of handling bursts of different bit-rates in the same transponder. These developments will be accompanied by increased emphasis on digital terrestrial communications probably involving a much larger variety of communications services than is available today.

References

1. P.P. Nuspl, K.E. Brown, W. Steenaart, B. Ghicopoulos, "Synchronization Methods for TDMA", Proceedings of IEEE, March 1977.
2. R.K. Kwan, "The Telesat TDMA System", International Conference on Communications, 1975, San Francisco, pp. 44-1 to 44-5.
3. "Statement of Work for Time-Division Multiple-Access Test Bed", INTELSAT, August 1976, Tech. Rep. RFP-15-836, Exhibit A.
4. W.G. Schmidt, "The Application of TDMA to the Intelsat IV Satellite Series", COMSAT Technical Review, Vol. 3, pp. 257-274, Fall 1973.
5. R. de Buda, "Synchronization for TDMA Experiments", CGE Technical Information Series RQ72EE4, June 1972.
6. P.P. Nuspl, R. de Buda, "TDMA Synchronization Algorithms", Conf. Rec. EASCON'74, Washington, D.C., October 1974.
7. P.P. Nuspl, N.G. Davies, R.L. Olsen, "Ranging and Synchronization Accuracies in a Regional TDMA Experiment", International Conference on Digital Satellite Communication - 3, 1975, Kyoto, Japan, pp. 292-300.
8. R. Gagliardi, W.C. Lindsey, and C.C. Weber, "Bibliography of Multiple Access Systems", University of Southern California, Interim Technical Report No. DOT-TSC-414-1, September 1972.

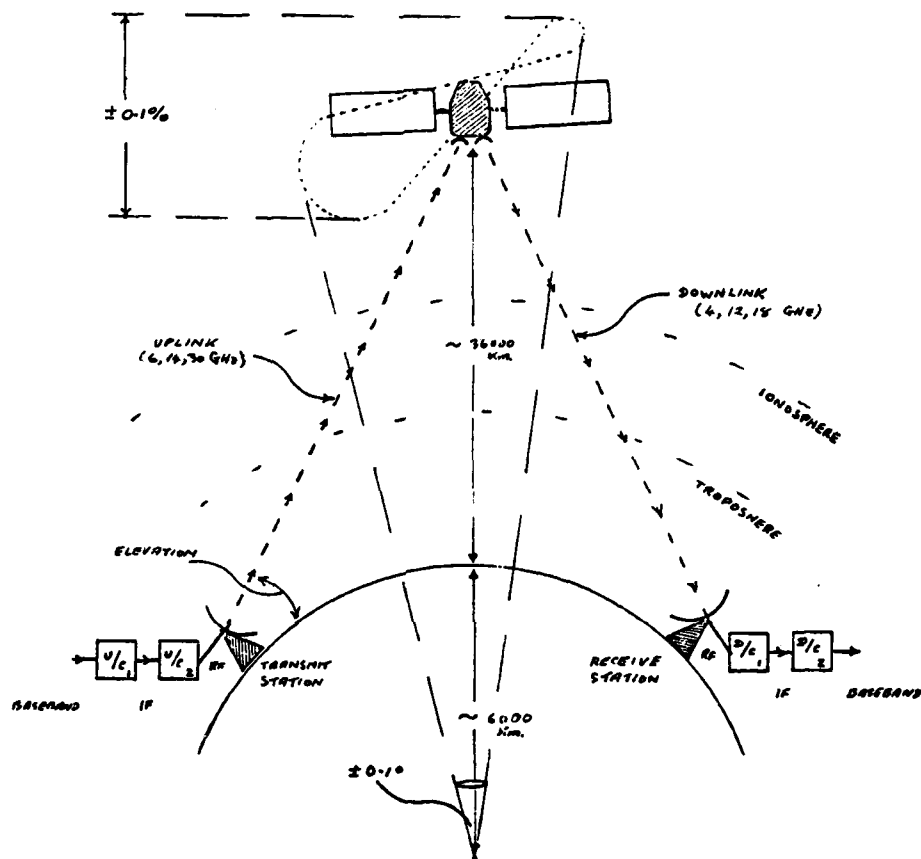


Figure 1: One Way Satellite Link
-- Geo-stationary Orbit

IDENTIFIER	ORGANIZATION	DATES	SATELLITE	CLASS OF SYNCHRONIZATION	PREAMBLE (# of symbols)	GUARD TIME (ns)	FRAME (μ s)	MODEM	RATES (Mb/s)	NETWORK
1. NATE	Intelsat	1966-68	IS-I	Reference Burst with Self-Ranging	82	220	125	2 ϕ CFSK	6.176	2 + 1 simulated
2. SNAX	ECL Japan	1968-71	ATS-1 ATS-3	Reference Burst, Clock Coherent	36	0	125	4 ϕ PSK	13.664	3
3. ITT	FDD Labs Japan	1969	IS-III	Reference Burst with Low-Level Ranging	38	< 100	125	4 ϕ DPSK	50	3 tested 10 design
4. DCLJ	COMSAT	1970	--	Low-Level Spread- Spectrum Coherent Detection	-- 18 + 1	100	832	4 ϕ CFSK	40 + 20 + 10	1 } design 7 } (20 max) 7 }
5. FRANCO- GERMAN	French-German Consortium	1973 1975	IS-III Symphonie	Reference Burst with Self-Ranging, Burst- Coherent (parallel) Detection	40	80	750	2 ϕ DPSK 4 ϕ DPSK	100/50	2 tested 30 design
6. ---	NTT, Japan	1975	30/20 GHz Dom. Sat.	Reference Burst, Clock Coherent	63 + 15	10 + 10	328	2 ϕ CFSK	100	10 design
7. Telesat TDMA	Telesat/ Teleglobe	1976	ANIK	Reference Burst	48	196	250	4 ϕ CFSK	61.248	2 operational
8. CENSAR	CRC/DOC Canada	1976-77	CTS	Centralized	31 + 4	< 30	125	2 ϕ DPSK	65 32.5	5 tested
9. ESA TDMA	ESTEC/ESA	1977	OTS	Low-Level M-Sequence		< 100		4 ϕ CFSK	60	
10. TDMA Test-Bed	Intelsat/ COMSAT	1976-	IS-V	Reference Burst + Self Lock, or Feedback, or Slave	72	~ 35	750	4 ϕ CFSK	120	60 design 4 to test

Figure 2 Partial Record of TDMA Activity Over Last Decade.

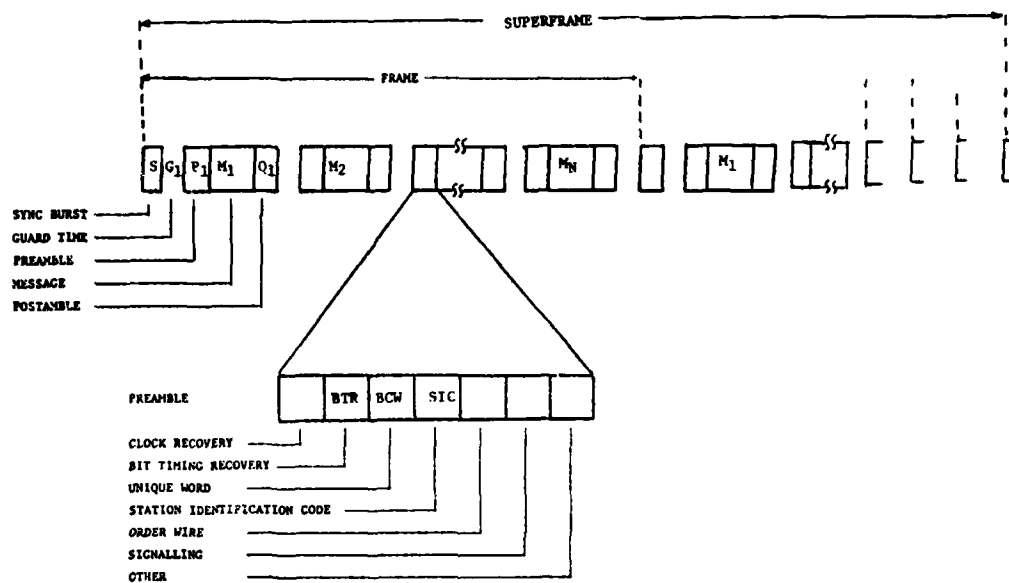


Figure 3 Typical Frame Format.

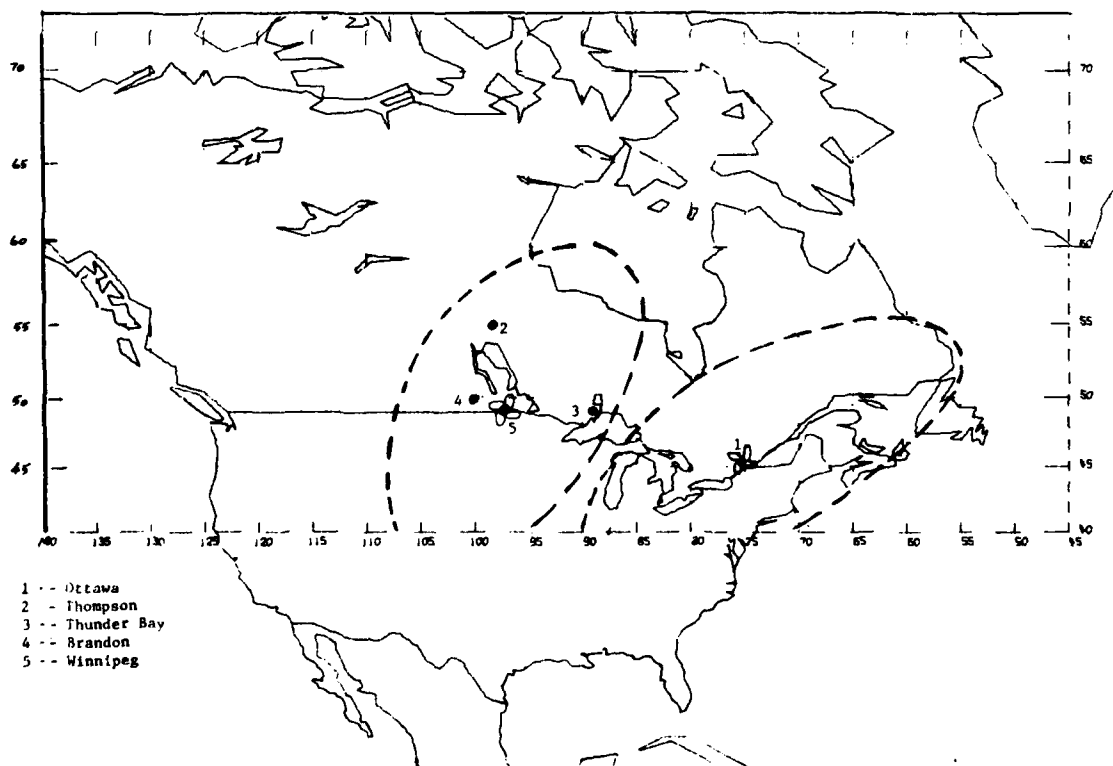


Figure 4 Area Coverage of CTS Spot Beam for Phase II of the CENSAR Experiment.

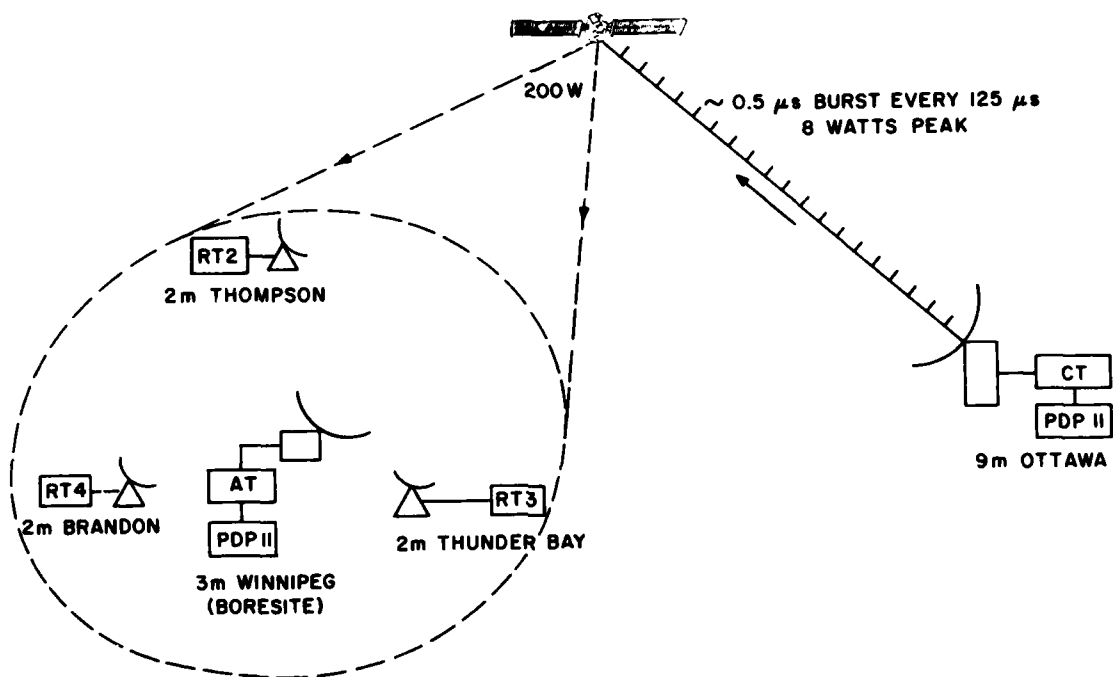


Figure 5: CENSAR Transmit.
Figure 5 CENSAR Transmit.

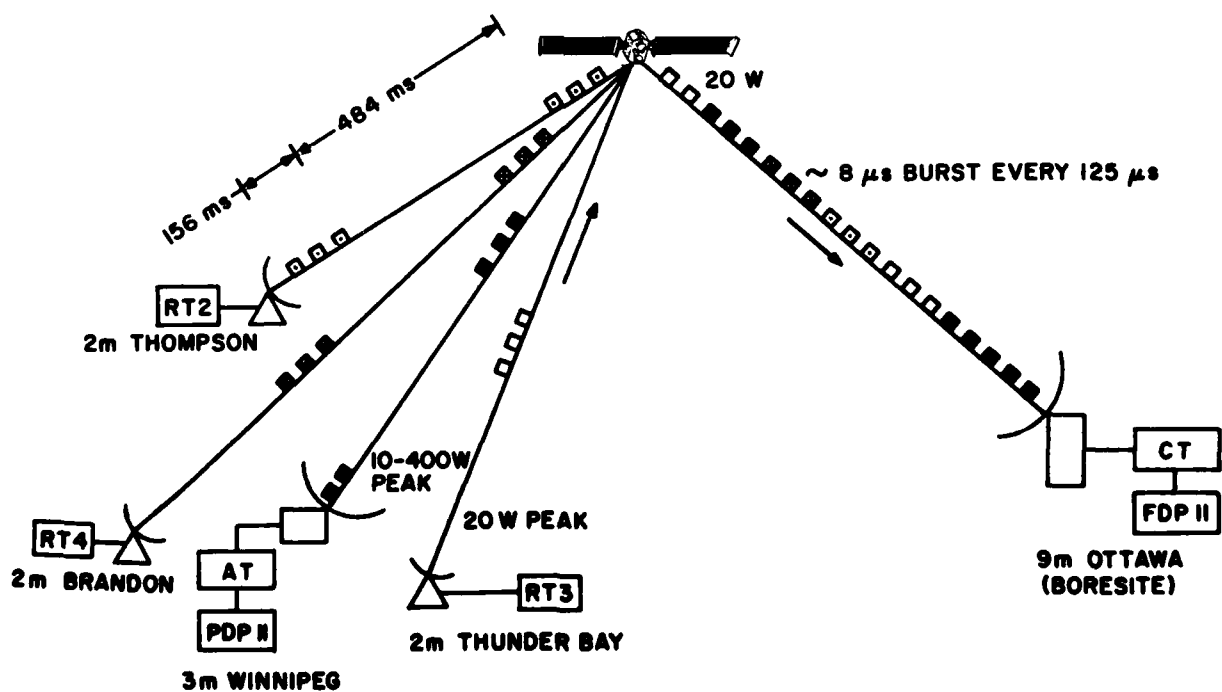


Figure 6 CENSAR Receive.

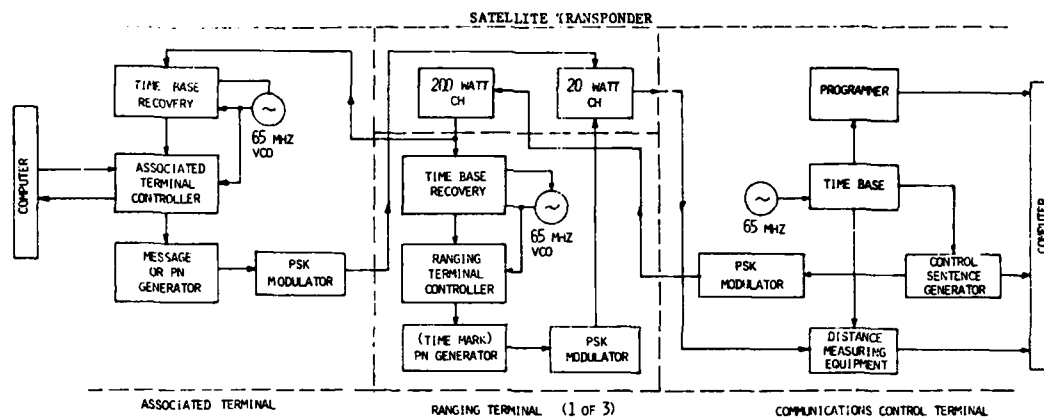


Figure 7 Block Diagram of CENSAR System.

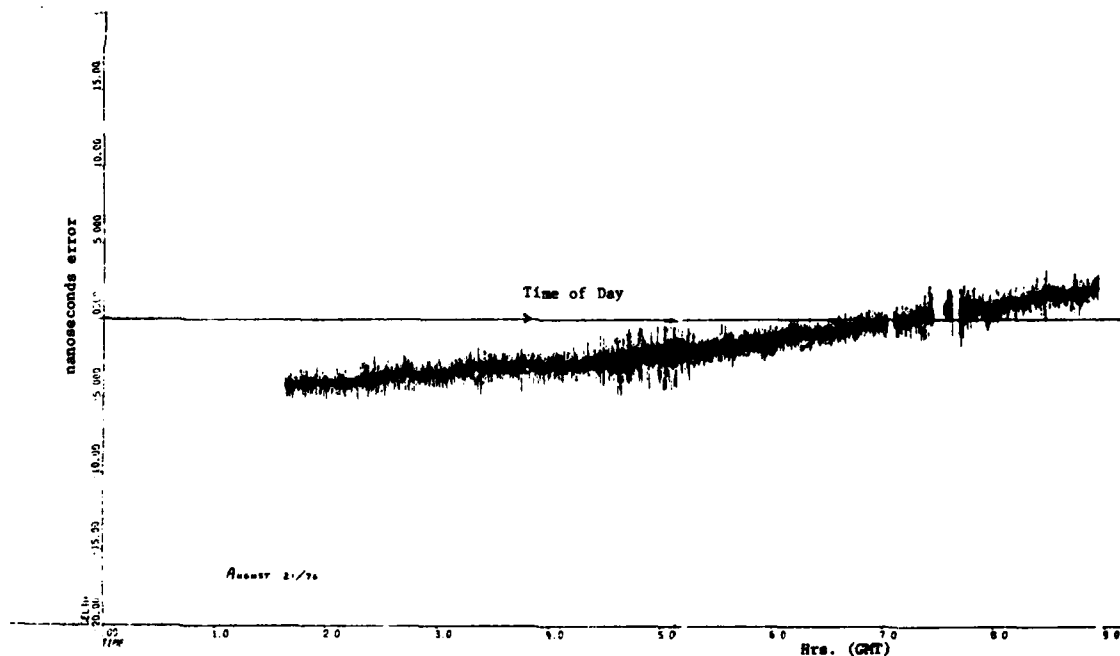


Figure 8: Error in Calculated Delay.

Figure 8 Error in Calculated Delay.

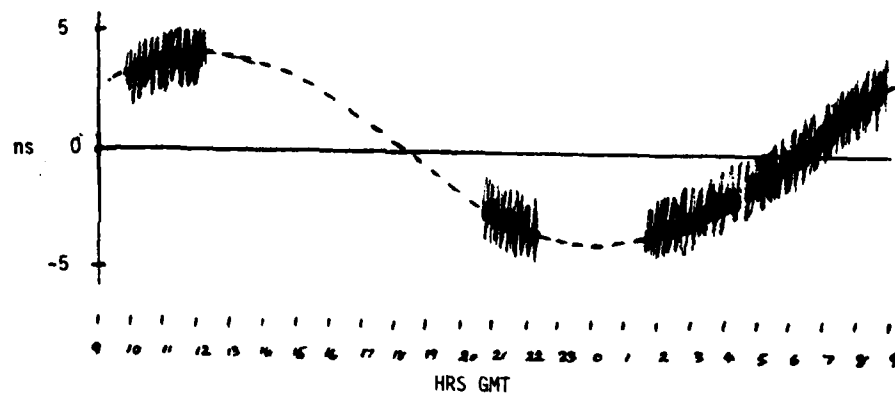
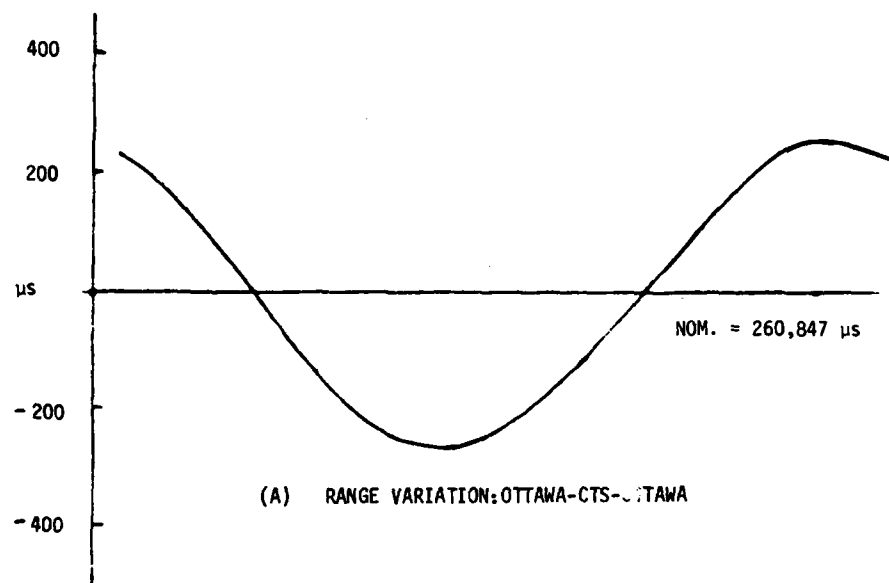


FIG. 9. RANGE VARIATION AND CORRESPONDING SYNCHRONIZATION ERROR, 20/21 AUGUST 1976.

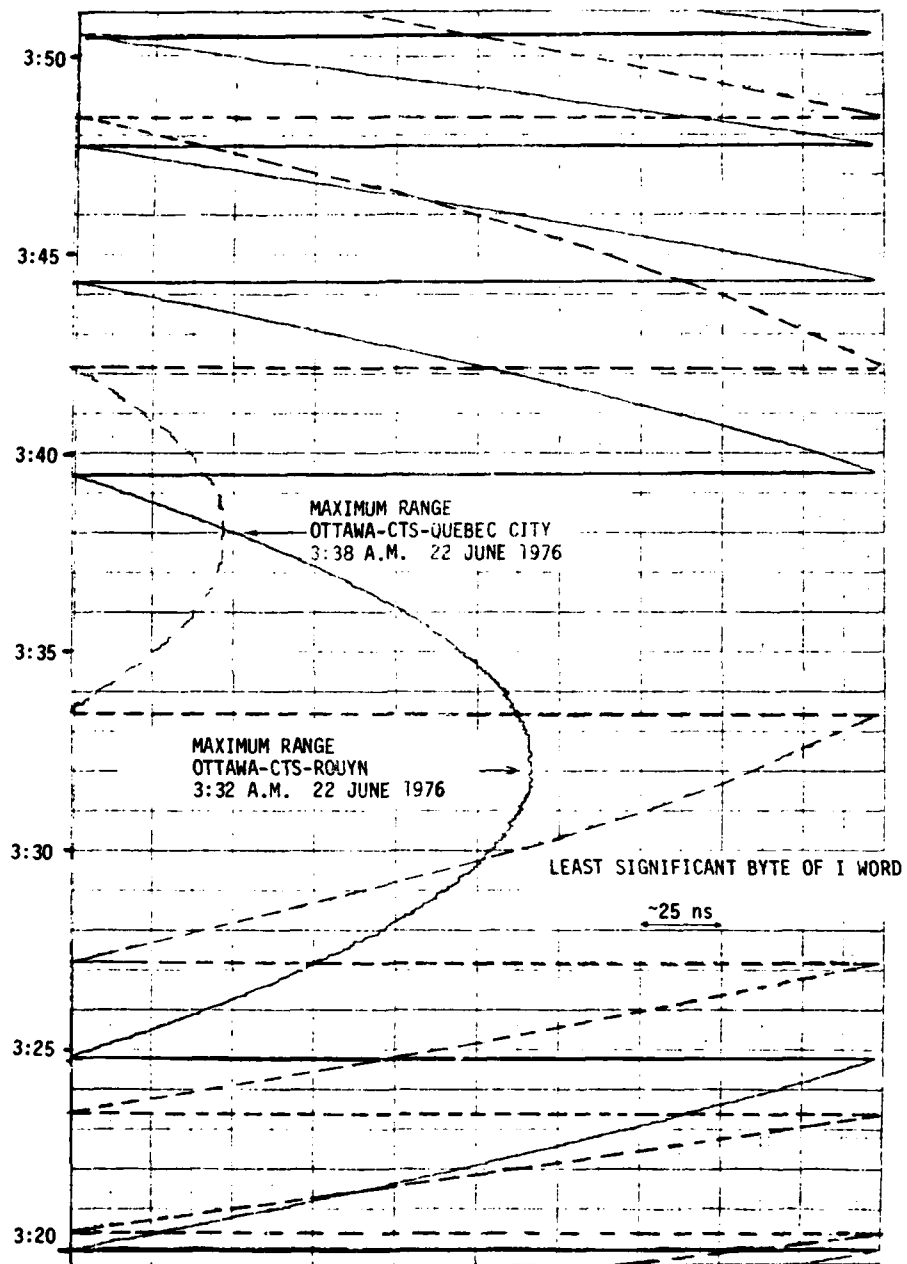


FIG. 10 . REAL TIME MEASUREMENTS OF THE RANGES TO TWO STATIONS.

FREQUENCY CONTROL AND TIMING REQUIREMENTS FOR COMMUNICATIONS SYSTEMS

Peter Kartaschoff, Swiss Post Office
Research and Development Division,
Bern, Switzerland *)

Summary

Recent evolutions of requirements for frequency control and timing in commercial communications systems are reviewed. The trend towards more stringent specifications is illustrated by means of two different examples: wideband analog transmission using frequency division multiplex (FDM) and digital networks using time division multiplex transmission (TDM) and time ordered switching. In both cases improved performance is required due to the impact of data communications.

In FDM systems the emphasis is on the phase noise spectral density $S_{\phi}(f)$ for f -values in the voice frequency range 300 to 3400 Hz. Typical values of 120 dB below 1 (radian)² are under discussion for individual signal sources.

In digital networks the emphasis is on timing properties of the network clocks and the establishment and maintenance of synchronism. In the design of extended networks the characteristics of the clocks interact with the properties of transmission paths, each introducing timing jitter and some systematic effects. Recent studies within the International Telegraph and Telephone Consultative Committee (CCITT) of the ITU have resulted in recommendations specifying maximum time interval error (MTIE) as a characteristic parameter for network clocks. Interconnection of synchronous networks on the international level is provided by means of plesiochronous operation with clocks having long term stabilities of 1 part in 10^{11} or better.

The required reliability of network timing is to be obtained by means of redundancy and proper network organization. The relations between frequency stability parameters developed by metrologists and those required by the systems designers are discussed. Some open questions point to the need of more dialogue and further common studies among the various interested specialists.

Introduction

The purpose of this paper is to review the requirements on frequency control and timing equipment used in present and coming generations of commercial communications systems. The emphasis is on the systems applications of recent developments in the field of time and frequency technology.

The characteristics required from sources generating stable and accurate frequencies and timing pulse sequences are determined by the properties of the system they serve.

Information bearing electrical signals are functions of time and have a spectrum. The dualism of time domain and frequency domain analysis appears quite naturally in the subject discussed here. Specific technical background information on Frequency and Time applications to communications systems can be found in two papers ^{2,3} published in 1972 in a special

issue of the IEEE Proceedings. It is thus possible to concentrate our attention to the evolution of systems requirements which occurred during the past five years. In our case this evolution can be characterized by the impact of the extensive and rapid development of data communications on commercial communications services, especially on the public telephone network. This network has sometimes been described as being the biggest automatic plant in the world. Its dimensions are indeed global and include the geostationary orbit of communications satellites. In some countries such as the United States, Sweden and Switzerland almost every household has a telephone. Data traffic is only a small percentage of telephone traffic now but growing fast and in view of the huge investment in the existing telephone system there is a great interest to use the voice facilities for data communications as well. Many ways are open to achieve this aim.

Two examples can illustrate the current situation regarding time and frequency applications in this field.

To operate a communications system such as the public telephone network two basic techniques are required:

Transmission and switching. Transmission carries information from point A to point B and vice versa by any means of wires or radio. Switching establishes the transmission link between the two points at the time and for the time required. Some combination of both techniques lead to the economically important concept of multiplexing, i.e. combining several individual circuits for transmission over one single link in a network. Two main techniques for multiplexing are known:

Analog signals can be translated in the frequency domain and combined to form channel groups on the frequency axis, forming a wideband signal. At the receiving end, the individual channels are sorted out (demultiplexed) by means of bandpass filters. This method is known as "frequency division multiplex" (FDM). Established for many years, it is still in use on most wideband trunk lines.

Digital signals are conveniently combined sequentially on the time axis, forming a time division multiplex (TDM) signal. Retrieval of the individual signals usually requires the knowledge of their location on the time axis. Thus, a problem of timing or synchronization arises which could only be avoided by providing each signal with a unique address identifying its destination (i.e. sending telegrams).

Both cases offer almost an infinite variety of possibilities. The worldwide telecommunications complex would not have become operational without the

*) Opinions expressed in this text are those of the author and not necessarily those of the Swiss PTT Enterprises.

long-time effort of international standardization. This task is in the hands of the International Telecommunications Union headquartered in Geneva, Switzerland.

FDM-System Requirements

The translation of voice frequency signals (300 + 3400 Hz) is performed by means of single-side-band suppressed carrier amplitude modulation. The combination of the channels is done in several steps² requiring various carrier frequencies. The highest capacity coaxial cable system currently in use has a bandwidth extending up to about 60 MHz. In long distance communications a signal may be submitted to a maximum of 70 modulation steps in cascade and be assigned to any channel in the band. The carrier phase and the phase of the signal component add directly at each modulator. Thus, any carrier signal frequency different from its nominal value shifts the frequencies of the signal components accordingly. For the purpose of maintaining the original pitch of transmitted voice and keeping the tolerances of audio frequency command signals within convenient tolerances a limit of ± 1 Hz departure from nominal is currently in use.² The highest carrier frequencies in a 60 MHz system must therefore be accurate to $\pm 1,7$ parts in 10^8 .

Obviously, it is very tempting to use the existing and continuously available telephone network for data transmission. This requires an interface unit (modem) between the data processing equipment on each subscribers side and the respective telephone lines.

The modem converts the digital data signal into a modulated audio carrier for transmission over the telephone circuit and vice versa. The audio bandwidth, level, signal to noise ratio and the allowed spectral characteristics are determined by international standards established by the Consultative Committee for Telephony and Telegraphy (CCITT) of the ITU.⁵ Work towards optimum use of these channels (originally designed for voice transmission and not for data transmission) has resulted in the use of phase-shift-keyed audio subcarrier signals. For transmission rates of a few kilobits per second four-phase (QPSK) or eight phase digital modulation is used. Typical examples are 2400 bits/s QPSK on ordinary switched circuits and up to 9600 bits/s on high quality leased circuits. A typical example is the 4800 bit/s modem specified in the CCITT recommendation V27.⁵ 8-phase PSK is used on a subcarrier of (1800 ± 1) Hz. Each of the possible eight phase transitions corresponds to a defined 3-bit word.

Obviously, phase jitter can lead to an erroneous detection of a phase transition. The maximum allowable phase tolerance is $\pm 22,5$ degrees, since the nominal eight phase angles differ by 45 degrees. Among the many possible causes of phase jitter that due to carrier generator phase noise should not be a dominant one. Current estimations⁶ of allowable random jitter lead to a margin of $\pm 7,5$ degrees peak-to-peak for carrier noise contribution. Assuming a normal distribution of the phase angle and a probability of exceeding the margin of 5×10^{-7} , the allowable RMS phase deviation is

$$\delta_{\phi_{\text{RMS}}} \leq \pm 1,5^\circ$$

It is further assumed that this jitter is a superposition of two approximately equivalent parts, namely that due to noise in the transmission channel and that due to the carrier generator. This leaves about $\pm 1^\circ$

RMS jitter to be allowed from the carrier generators.

We then may distribute this equally over the 70 generators assumed on a long distance circuit. The noise from each is not correlated to that of the others, at least in the audio band from 300 to 3400 Hz. We thus obtain for each single generator:

$$\delta_{\phi_{\text{RMS}}} \leq \pm \frac{1}{\sqrt{70}} = \pm 0,18^\circ \text{ RMS}$$

The phase jitter must not exceed this limit at the highest carrier frequency used, i.e. at 60 MHz.

It is now easy to compute the phase noise spectral density. Assuming white noise within the bandwidth B we have:

$$\langle \delta_{\phi}^2 \rangle = \delta_{\phi_{\text{RMS}}}^2 = \int_{f_1}^{f_2} S_{\phi}(f) df = B \cdot S_{\phi} \quad (1)$$

$$\text{with } \delta_{\phi_{\text{RMS}}} = 0,18^\circ = 3,14 \cdot 10^{-3} \text{ radian}$$

$B = 3100$ Hz, we obtain

$$S_{\phi} = \frac{9,87 \times 10^{-6}}{3100} = 3,184 \times 10^{-9} \text{ (radian)}^2/\text{Hz} \quad (\text{at } 60 \text{ MHz})$$

In current practice, the carrier generator equipment is driven by a master frequency supply system which generates a set of basic frequencies, e.g. 4 - 12 - 124 - 440 - 2200 kHz.

The layout of the frequency synthesis process implies frequency multiplication from 2200 kHz up to about 60 MHz plus addition or subtraction of multiples or submultiples of the other values. It is well known that frequency multiplication by a factor n multiplies any phase fluctuations by the same factor and thus the phase noise spectral density by n^2 .

Even in assuming an ideal frequency multiplier from 2200 kHz up to about 60 MHz, we must require from the basic 2200 kHz generator a spectral density of

$$S_{\phi} = \left(\frac{2,2}{60}\right)^2 \times 3,184 \times 10^{-9} = 4,28 \times 10^{-12} \text{ (radian)}^2/\text{Hz at } 2200 \text{ kHz}$$

Expressing this value in decibels with reference to $0\text{ dB} = 1 \text{ (radian)}^2/\text{Hz}$, this equal to

$$S_{\phi} = 10 \log (4,28 \times 10^{-12}) = -113,7 \text{ dB at } 2200 \text{ kHz}$$

The purpose of this example is to give an idea as regards the orders of magnitude required. The example is conservative in assuming the worst case of having all transmission sections on carrier frequencies near 60 MHz. On the other hand, no margin has yet been included for frequency multiplier and mixer noise in the carrier synthesis process.

For the development of new carrier generating equipment the Swiss Post Office issued a provisional specification given in Fig. 1. It includes a flat portion from 50 to $(4000 - \epsilon)$ Hz of -117 dB. At 4 kHz and above, this is relaxed to -83 dB. Below

50 Hz, the limit is relaxed according to the state of the art in crystal oscillators, proportional to f^{-1} down to 1 Hz and f^{-3} below 1 Hz. Test on prototype equipment have shown that this is feasible.

TDM System Requirements

As long as we deal only with point to point digital links, the synchronization problem is solved by techniques known for many years. Thousands of PCM systems are in operation throughout the world transmitting 24 voice channels at 1544 kbit/s (in the USA, Canada and Japan) or 30 channels at 2048 kbit/s (in Europe) over a single pair of wires. In many cases this technique has given a second life to cables laid 40 years ago or more.

A different situation has evolved with the development of electronic switching techniques and entirely digital communications networks. Again, some distinction is to be made between networks devoted exclusively to telephone traffic and those involving data communications, due to the fact that digitized voice (e.g. PCM) is considerably more tolerant to occasional bit errors than data - at least if we wish to exploit in the most effective way the inherent high data transmission capacity of a switched digital network.

Whatever the size and complexity of such a network - conceptually there does not seem to be a limit - the question about synchronism is present by the mere fact that digitized information is time ordered. The degree of synchronism required depends on the parameters and organization of the network. A slight lack of synchronism leads either to the loss or to the duplication of some information elements. However, imperfect synchronism is not the only source of errors in such systems. Therefore, even perfect synchronization - probably not feasible in a geographically extended system - would not solve all the problems.

Only very small systems can rely on a single clock generator. Its failure shuts down the system. In any network above the minimum size there are many clock generators. Synchronization always requires in principle the fulfillment of two conditions:

1. The clock rates (frequencies) must be equal within a specified tolerance.
2. Switching operations, i.e. logical transitions must occur at specific instants.

Two main factors act against the preservation of synchronism:

1. Clock imperfections (adjustment error, drift)
2. Transmission delay variations.

At the receiving end of a transmission link it is not possible to distinguish between the causes of either effect, only their superposition can be seen. Transmission delay variations always remain within some bounds and buffer memories help to take care of them. The incoming data stream is written into the memory at its own rate and read out at the rate of the local clock. Obviously, this involves additional delay. If the local clock is slow, the buffer overflows and information is lost. Synchronization, i.e. establishment and maintenance of synchronism is obtained by acting on the clocks and on the delays. Bit stuffing (justification) techniques acting on the delays only are used on high capacity (high bit rate) point-to-point transmission links. Exclusive use of stuffing allowing asynchronous

clock operation does not appear to be the most economic solution for a large switched network. Many synchronization schemes have been described and proposed during the past 15 years.

In these proposals we can distinguish two main basic concepts of the way in which the clocks are controlled. One could be called "democratic" since it is based on mutual control of all oscillators having the same rank in the network, each taking its reference from the average of a number of incoming signal streams^{8, 9, 10, 11, 12}. The other method is of more hierarchical nature, i.e. there is a master clock which controls other oscillators of lower rank^{6, 12, 13}. In mutual control (frequency averaging) systems the frequencies of the oscillators converge towards their average value on the long run. However this average value is not constant if all clocks have the same weight and some show drift (aging) as this will be the case if crystal oscillators are used.

Introduction of oscillators stabilized to an atomic reference implies usually some form of hierarchization, however, there could still be some forms of mutual control should there be a failure of the atomic reference.

Performance, reliability, cost and flexibility for future growth are parameters to enter in the planning and design process. There is a wealth of possible and a number of good solutions but no single specific method has yet clearly emerged as being the best one.

Discussions held during the last five years in the framework of CCITT Study Group XVIII (formerly Special Study Group D) have led to a recommendation which constitutes the level of agreement reached in 1976. It leaves enough freedom for the organization of national networks but sets minimum standards for the interfaces on the international level. It excludes totally asynchronous networks but makes a clear distinction between two classes of synchronous networks:

Networks in which the clocks are controlled by means of one of the synchronization schemes mentioned above are called synchronous. Networks in which the master oscillators are not continuously controlled but adjusted within close margins and of such high stability that there are only very occasional slips are called plesiochronous (from greek: plesios = near, i.e. nearly synchronous). On the international level, plesiochronous operation is being envisaged at least for an initial period, since a binding agreement on an unique control scheme is extremely difficult to achieve.

The general condition for the international network requires that the rate of occurrence of slips in any 64 kbit/s channel should not be greater than one in every 70 days per international exchange office. Furthermore, where a national network or sub-network is controlled by a single reference clock the frequency of the clock must not be affected by control signals generated within the national network or sub-network. Similar conditions apply to the case of mutual synchronization between reference clocks.

Detailed specifications including clock performance are given in the form of allowable time interval error (TIE). A time interval error is the difference between a nominal interval separating two instants generated by the clock and the true time interval as determined by the international standard time scale. The recommended performance of the output digit stream from an exchange directly controlled by a reference clock is given as follows:

1. The TIE over any period up to 2^{11} unit intervals shall not exceed $1/8$ of an unit interval. (Equivalent frequency inaccuracy of 6×10^{-5})
2. The TIE over any period of 2 seconds shall not exceed 200 ns (Equivalent frequency inaccuracy of 2×10^{-7})
3. The TIE over any period of 50 seconds shall not exceed 500 ns (Equivalent frequency inaccuracy of 1×10^{-8})
4. The TIE over any period of 1000 seconds shall not exceed 1000 ns (Equivalent frequency inaccuracy of 1×10^{-9})
5. Over any period of S seconds, the TIE shall not exceed $(10^{-2} S + 20000)$ ns. (Equivalent frequency inaccuracy of up to 1×10^{-11} plus a maximum peak - to peak jitter of 20000 ns)

Note 1: 2^{11} unit intervals is 1 ms at 2048 kbit/s and approximately 1,3 ms at 1544 kbit/s

Note 2: 20000 ns is approximately 40 unit intervals at 2048 kbit/s and 30 unit intervals at 1544 kbit/s.

It is important to keep in mind that this specification regards not only the reference clock but the whole equipment constituting the international interface. In particular, it comprises some margin to allow short term time interval departures required for the operation of synchronization systems. The term "shall not exceed" should be completed by a statement of (very low) probability of violation.

The time interval error as defined above can be related to the clock oscillator parameters used in the field of time/frequency metrology as follows:

If the clock's normalized frequency departure ("fractional frequency offset") $y(t)$ defined in the literature ^{15, 16} is known, the TIE accumulated between the instants t_k and $t_k + \tau$ on the reference time axis is given by

$$x(t_k, \tau) = \int_{t_k}^{t_k + \tau} y(t) dt = x(t_k + \tau) - x(t_k) \quad (2)$$

On the other hand, any measurement of $y(t)$ cannot be made directly, but only in the form of samples averaged over a finite time interval τ and this is:

$$\bar{y}(t_k, \tau) = \frac{1}{\tau} x(t_k, \tau) \quad (3)$$

Repeated measurements of this quantity lead quite naturally to the computation of the average TIE and this is simply related to the two sample (Allan) variance

$$\sigma_y^2(\tau) = \frac{1}{2\tau^2} \langle (x(t_k + \tau, \tau) - x(t_k, \tau))^2 \rangle \quad (4)$$

We then may write: ¹⁵

$$\sigma_x^2(\tau) = 2\tau^2 \sigma_y^2(\tau) = \tau^2 D_x^{(2)}(\tau) \quad (5)$$

where

$$D_x^{(2)}(\tau) = \frac{D_\phi^{(2)}(\tau)}{(2\pi\nu_0\tau)^2} \quad (5a)$$

using the relation between the two sample variance and the second phase structure function introduced by Lindsey and Chie ¹⁶. To the extent in which it is allowed to predict future behaviour from past performance we may use $\sigma_x(\tau)$ as a predicted RMS TIE,

assuming that the TIE samples are normally distributed.

The last assumption calls for some caution: Frequency drift is to be excluded, i.e. its effects should be small compared to the random fluctuations. Furthermore, as the specification comprises not only clock errors but also timing variations due to other system characteristics, it would not be wise to use up all the allowed margin by clock instabilities.

Fig. 2 gives an idea about the possibilities to accommodate state of the art clock performance within the recommended specifications. The solid line reproduces the TIE limits recommended in CCITT G.811. The dashed line (1) shows $\sigma_x(\tau)$ derived by using eq. (5) on published manufacturer's specifications of typical commercial cesium stabilized oscillators. The dashed line (2) shows an example of a possible failure mode of the reference clock: removal of the cesium reference and continuing operation with a good crystal oscillator which ages 3 parts in 10^{11} per day. The TIE specification would be violated after 53 hours. This leaves enough time for replacement of the defective unit. It is however an open question whether all or part of the allowed 20 microsecond margin can be used in such a case.

Since a total failure of the timing system leads to the disruption of high traffic capacities, maximum reliability is at least as important than specified timing performance. For a typical international exchange the required mean time between total system failures should be at least of the order of 50 years supposing normal preventive maintenance. There is still some discussion about how the reliability should best be defined but the figure cited appears realistic on the basis of past experience with FDM carrier frequency generators operating in the author's country.

Proper network organization is a major factor for achieving high operational reliability. A large system must not depend exclusively on one single timing center but split up into regions. These must be capable to remain operational despite failures affecting adjacent regions. Regional centers will be equipped with master oscillator supplies comprising at least two redundant units. ¹⁸ In Switzerland the long distance network comprises about 50 regional centers having each its carrier generator supply for FDM systems. The future TDM systems will of course be integrated into the existing facilities. There is no firm commitment yet as to how many of the 50 stations will be designated as reference stations with cesium stabilized master oscillators but there will be more than one. All centers are already linked by a reference frequency supply system which is currently being reevaluated for modernization.

One area requiring more study is the remote control of reference frequency generators via digital links. Some recent results of jitter measurements on 2048 kbit/s lines containing up to several hundred regenerating repeaters in cascade are available ¹⁹ but there is still a need of more experimental data on various systems.

Conclusion:

The review of work published during the past fifteen years on the subject of time and frequency and communications technologies seems to indicate that considerable work has been done in parallel without very much dialogue between specialists in the two fields. Now this situation appears to have

changed considerably and the examples have been discussed here as an attempt to bring the two disciplines more closely together. Much work remains to be done and this points to an exciting common future.

Acknowledgements:

The author wishes to express his gratitude to Mr. Willy Kleir, Director of the Swiss PTT Research and Development Division for the permission to publish and to Dr. Walter Neu, Hans K. Pfyffer and Kurt Hilty for many helpful discussions preceding the preparation of this paper.

References

- (1) R.B. Blackman and J.W. Tukey "The measurement of power spectra", Dover New York, 1958
- (2) W.L. Smith, "Frequency and Time in Communications", Proc IEEE Vol. 60, No 5 pp 589-593 (May 1972)
- (3) J.W. Pan, "Synchronization and Multiplexing in a Digital Communications Network", Proc IEEE Vol. 60, No 5. pp 594-601 (May 1972)
- (5) CCITT Green Books, ITU Geneva (reedited every 4 years, next edition scheduled summer 1977), 9 volumes
- (6) H.K. Pfyffer, Swiss PTT Report No VL 2.013A, October 21, 1974 (unpublished)
- (7) J.R. Mensch, Future DCA Objectives in Communications Network Timing and Synchronization, Proc. 5th PTT/Planning Meeting, NASA X-814-225, Goddard Space Flight Center, Greenbelt MD 20771 (December 1973).
- (8) J.R. Pierce, "Synchronizing Digital Networks", Bell System Technical Journal, Vol 48, pp 615-636 March 1969
- (9) H.A. Stover, "A Time Reference Distribution Concept for a Time Division Communication Network, Proc. 5th PTT/ Planning Meeting, NASA X-814-225, Goddard Space Flight Center, Greenbelt MD20771 (December 1973)
- (10) W.T. Duerdoh, "A Method of Synchronizing a Network of Digital Exchanges, Proc. International Switching Symposium 1972, p. 36
- (11) H. L. Hartmann, "Das Synchronisierproblem in digital arbeitenden Informationsnetzen, Habilitationsschrift Tu München, Braunschweig 1972
- (12) R.G. DeWitt, "Network Synchronization Plan for the Western Union All Digital Network", Telecommunications, p. 25, July 1973
- (13) R. H. Bittel, W.B. Elsner, H. Helm, R. Mukundan and D.A. Perreault, "Clock Synchronization Through Discrete Control Correction", IEEE Trans. on Communications p. 836 June 1974.
- (14) CCITT Recommendation G. 811 (Geneva, 1976)
- (15) NBS Monograph 140, "Time and Frequency" U.S. Govt. Printing Office, SD Catalog No. C13. 44:140 (May 1974)
- (16) P. Kartaschoff, Frequency and Time, Academic Press Inc. (London, New York and Sydney), 1977 (in press)
- (17) W.C. Lindsey and C.M. Chie, "Theory of Oscillator Instability Based upon Structure Functions, Proc. IEEE, Vol. 64, No 12, pp 1652-1666 (December 1976)
- (18) E.P. Graf und F. Walther, "Generation of base-band frequencies for FDM and TDM Telecommunications", elsewhere in these proceedings
- (19) A. Kaeser, "Quelques mesures de taux d'erreur et de gigue en ligne sur les systèmes à 2 Mbit/s sur câbles", Bulletin Technique PTT (Suisse), Vol. 54, pp. 428-430 (November 1976)

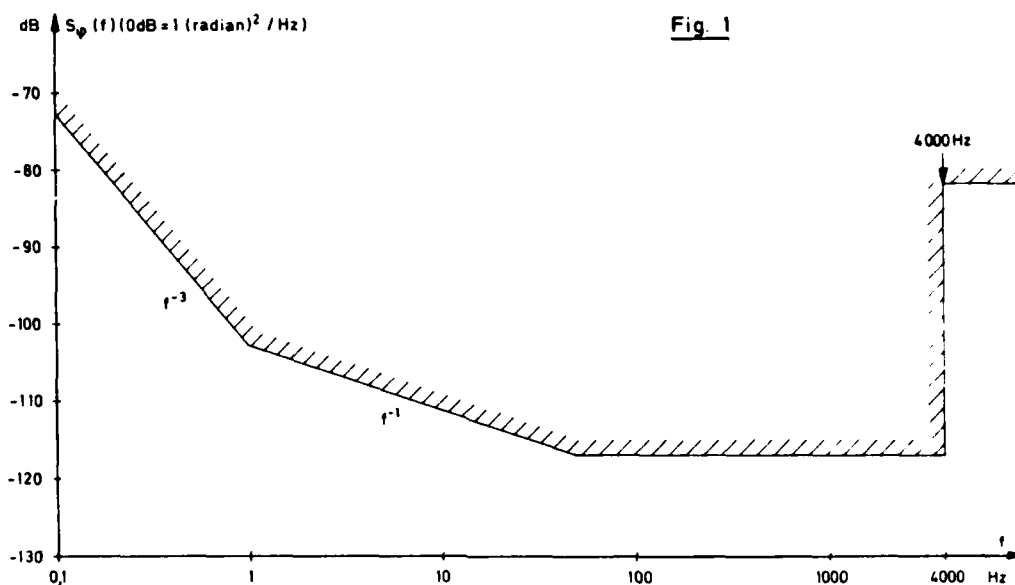


Fig. 1 Provisional Carrier Generator Phase Noise Spectral Density Specification, Swiss Post Office 1975

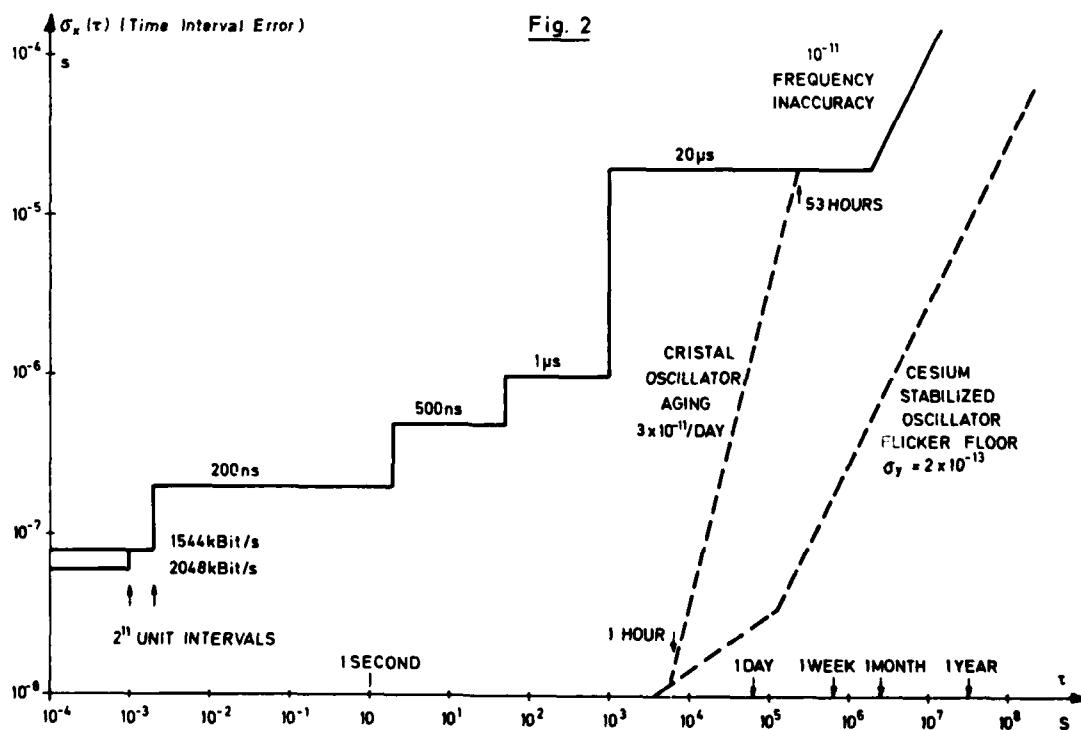


Fig. 2 TDM System Time Interval Error, Solid line: CCIT Recommendation G 811 (1976) Dashed lines: State of the art oscillator TIE prediction (see text)

GENERATION OF BASE-BAND FREQUENCIES FOR FDM AND TDM TELECOMMUNICATIONS

Erhard P. Graf and
Bernard Walther
Oscilloquartz S.A.
Neuchâtel, Switzerland

Summary

The introduction of new techniques in telecommunications has created particularly demanding specifications on the quality of base-band frequencies. The long term ($\tau > 100$ s) stability of the master clock must be less than 1×10^{-11} . Moreover, the different base-band frequencies must have a spectral purity for Fourier frequencies from 100 Hz to 4 kHz which can only be achieved by high-quality quartz crystal oscillators.

A new development which meets these requirements as well as those of reliability and redundancy demanded by telecommunications systems is described.

Introduction

The ensemble of base-band frequency generators must supply the needs of a communications center with 6 FDM and 3 TDM frequencies, according to very high specifications. A modular construction permits equipping the frequency generating rack in a partial or complete fashion, to synchronize the signals by the distribution of an external reference frequency or to generate the reference frequency locally with the aid of a cesium standard.

Each output frequency is derived from a quartz oscillator, contained in the generator synthesizer module, and each module provides either 15 or 25 isolated outputs according to the frequency. For reasons of redundancy, there are two generator-synthesizer modules for each frequency; all units are in turn servoed to a 5 MHz internal reference. Thus a completely equipped rack will have a total of 420 outputs as follows:

2 x 15 outputs at: 440 kHz
2.108 MHz
2.2 MHz
2.048 MHz

2 x 25 outputs at: 4 kHz
12 kHz
124 kHz
4.096 MHz
16.896 MHz
5 MHz

Dual power-supplies connected by "or" diodes as well as a monitoring system of urgent and non-urgent alarm signals serves to obtain a high reliability and ease of maintenance.

Block diagram

Fig. 1. Starting with the cesium frequency standard, which can be replaced by an external reference, one recognizes first the system of generation of the internal reference. The 5 MHz frequency is the backbone of the rack, if it is missing, all synthesized outputs are cut off. The present system assures a high reliability, since it is necessary at any point to have two simultaneous failures to cause a general interruption.

By reason of power-combiner and power-splitter elements which have 50 Ω impedance, all distribution of 5 MHz is at 50 Ω .

Three units of complimentary character joined together by a vertical servo structure combine to form a reference with the characteristics:

- a) spectral purity of a quartz oscillator ≤ -135 dB for Fourier frequencies > 10 Hz
- b) short term stability $\sigma_T, \tau = 1$ s,
 3×10^{-12}
- c) long term stability of a cesium standard $< 5 \times 10^{-12}$

The cesium standard determines the long term stability. By means of a combined analog and digital servo the cesium stability is given to the primary B-5400 quartz oscillator. This oscillator in turn servoed the second B-5400 oscillator by means of an identical loop. The servo time constant, about 130 s, is chosen such that the cesium standard does not degrade the quartz oscillator performance in the range where the latter has a better frequency stability.

In case of the absence or failure of the cesium standard, oscillator 1 continues to servo oscillator 2, and the frequency is determined by the B-5400 which has an aging rate of less than 1×10^{-10} per day, typically 5×10^{-11} per day. In the racks for FDM frequencies without cesium standards, two functional modes are possible for oscillator 1:

- a) the frequency control voltage is cut off, frequency is set manually.
- b) the servo loop memory is blocked and provides a fixed voltage to the oscillator control. At turn-on this voltage is at the center of control range.

the phase-locked-loop is in position of waiting : the frequency can be periodically readjusted by connecting an external standard. The state after the last correction is memorized.

The servo control assemblies and oscillators are identical, therefore, interchangeable. The units are adjusted in such a manner that the output signal and the reference are in phase.

Frequencies are synthesized by classic PLL, composed of a linear phase comparator, a quartz voltage controlled oscillator non-temperature controlled, frequency dividers and output amplifiers.

Cesium frequency standard

Fig. 2

The frequency standard is a model 3000, described at the Frequency Control Symposium in 1975.

The front panel controls concern routine measurements of equipment operation which are made a few days after placing the standard in operation, and then at regular intervals, 1 to 3 months. Other operating controls are accessible after removing the cover panel.

5 MHz phase locked loop

The high quality temperature controlled quartz oscillator is one of the most stable frequency sources available on the market today for observation times up to some hundreds of seconds. The slow drift of frequency necessitates a recalibration from time-to-time. The frequency instabilities of a cesium standard are smaller than those of quartz oscillators for measurement times in the order of hundreds of seconds.

Consequently, the servo time constant for locking a crystal oscillator to a cesium standard should be in the order of 100 seconds to take maximum advantage of the two frequency sources.

A proportional feed-back control system capable of maintaining the oscillation of a crystal oscillator in phase with the oscillation of a cesium standard must have considerable gain to be able to correct the slow frequency change due to aging of the quartz. However, it is extremely difficult to realize such a proportional servo loop having at the same time the necessary gain and a time constant of the order of 100 seconds.

On the other hand, a digital servo loop does not allow regulation of the crystal oscillator in steps sufficiently small such that the deviation in phase between the quartz oscillator and the cesium standard, will always be within desired limits. High capacity digital memories with output of 12 bits and conversion to analog voltage, for

example, permit a regulation of 4096 steps. This is insufficient to control the crystal oscillator for years and obtain at the same time a regulation sufficiently fine.

The present method is characterized by two feed-back control loops connected in parallel; one has digital configuration with an up-down counter memory, the other has an integral-proportional control configuration with an amplifier integrator. One obtains the optimum feed-back control conditions by the combined action of the two loops. The digital loop allows correction of the drift of the crystal oscillator in steps relatively large over a wide range, while at the same time the analog control loop provides a fine correction between the steps of the digital loop.

In case of absence of the standard frequency, the counter - memory of the digital loop retains the frequency of the crystal oscillator at the last value attained. On the other hand, as the proportional control loop requires a gain only sufficient for phase regulation between two steps of the digital loop, it is possible to realize an analog loop in which the time constant is sufficiently long to take advantage optimally of the characteristics of the two frequency sources.

A system block diagram is given in fig. 3, showing the main functions.

This system has two basic operating modes : the initialization mode which is necessary at the time of establishment of the reference frequency to quickly correct the frequency of the oscillator; the normal operation mode when only small corrections are necessary to compensate for the aging of the crystal. The control circuitry automatically cycles the system from one mode to the other, when necessary.

Operation of the servo system can be understood as follows:

There are four phases possible:

- 1) Phase I, initialization, which is necessary when establishing the control voltage, when connecting the reference frequency or when an accidental phase or frequency jump occurs.
- 2) Phase II, test, which permits automatic passage from phase I to phase III on resetting the integrator to zero.
- 3) Phase III, normal operation, in which phase the voltage derived from the Digital-to-analog converter progresses by steps each time the integrator output voltage reaches its limits.
- 4) Phase IV, blocked, in which the oscillator is free running because of lack of reference frequency, and the oscillator frequency is established by

the last value memorized by the D-A converter.

Phase I can also be manually started, only the analog servo mode is operational. Amplifier gain is 4 and the servo time constant is 5 seconds in order to increase the pull-in range. This phase permits a rapid acquisition and memorization of the oscillator frequency control voltage by measuring the phase difference at the reference frequency. The total range of the memory is swept in 16 seconds.

Phase III, normal operation, establishes a time constant of 100 s. This is accomplished by reducing the system gain by a factor of 400 and at the same time placing the digital and analog servo systems in parallel. One step of correction of the D-A converter corresponds to 5×10^{-11} parts change in frequency of the oscillator. The steps of correction of the voltage control by the D-A converter are only possible in an interval of 64 s so reducing the transmission of reference oscillator noise. An increase or decrease of the memory is therefore established by two criteria: integrator output voltage falling outside the comparator window reference voltage and, simultaneously, presence of correction pulses which follow at an interval of 64 s. The comparator window functions at a level 100 times greater than a voltage unit of correction, therefore, at a level which poses no drift-problems. To avoid transient effects, a correction step is always accompanied by resetting the integrator to zero. A typical variation with time of the control output voltages of each control-loop is shown in fig. 4

Frequency synthesis

Schematic block diagram fig. 5 shows the generation of frequency cited above by direct PLL synthesis. Considerable development effort was devoted to suppression of side bands at the comparison frequency and its harmonics, and to phase-noise spectral purity close to the carrier.

Results

Fig. 6 shows spectral characteristics of the internal 5 MHz reference frequency which corresponds to the performance of a non-perturbed oscillator.

Fig. 7 shows a similar measurement made on one of the synthesizer outputs.

Modular construction is illustrated in fig. 8, 9, 10 and 11.

Acknowledgments

The authors wish acknowledge the important contributions of C. Herbelin, A. Waefler, and J.-P. Wiedmer to the development of this equipment.

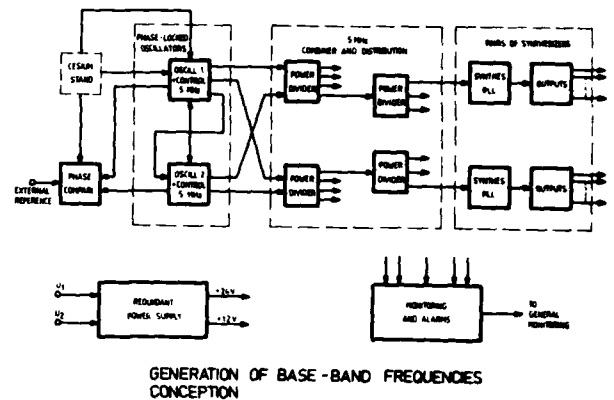


Fig. 1

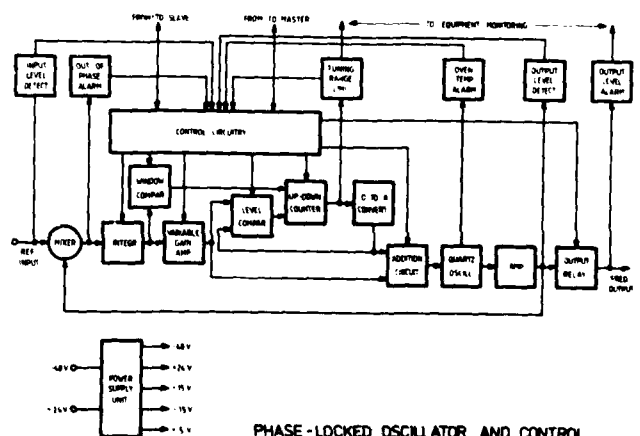
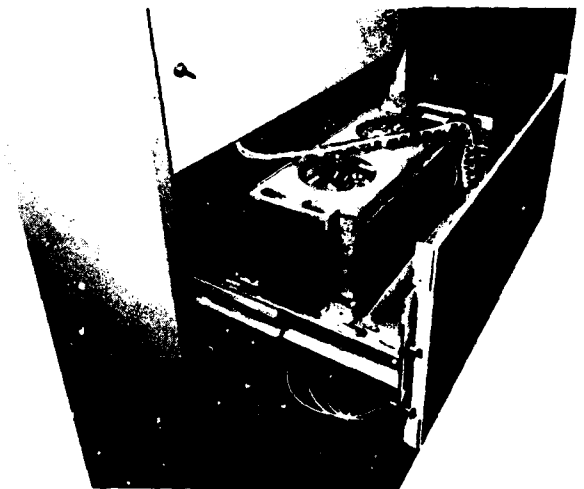
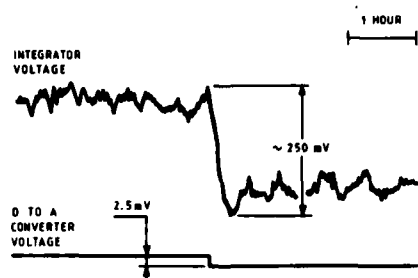
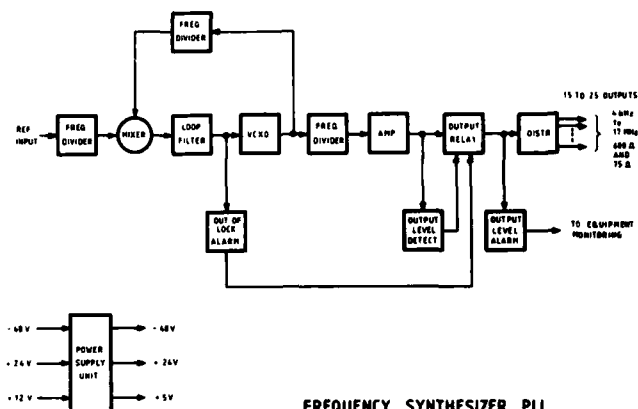


Fig. 3



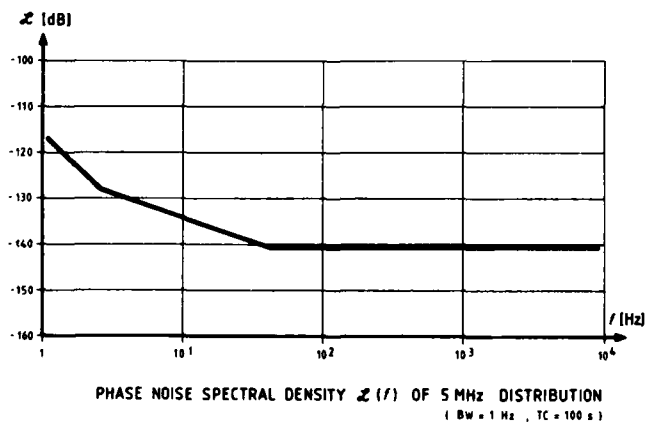
OSCILLATOR CONTROL VOLTAGES

Fig. 4



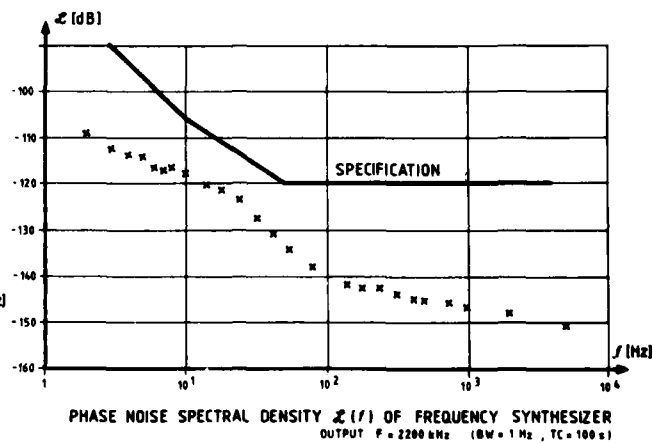
FREQUENCY SYNTHESIZER PLL

Fig. 5



PHASE NOISE SPECTRAL DENSITY $\mathcal{L}(f)$ OF 5 MHz DISTRIBUTION
(BW = 1 Hz, TC = 100 s)

Fig. 6



PHASE NOISE SPECTRAL DENSITY $\mathcal{L}(f)$ OF FREQUENCY SYNTHESIZER
OUTPUT $F = 2200$ kHz (BW = 1 Hz, TC = 100 s)

Fig. 7

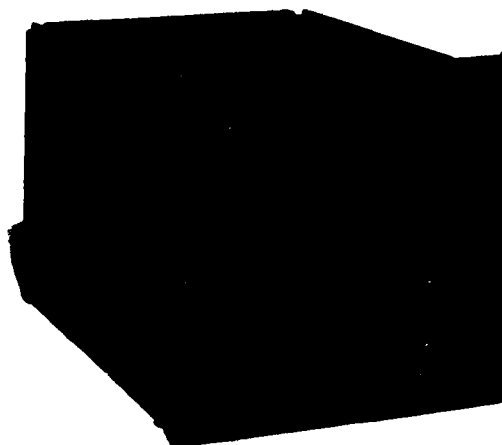


Fig. 8

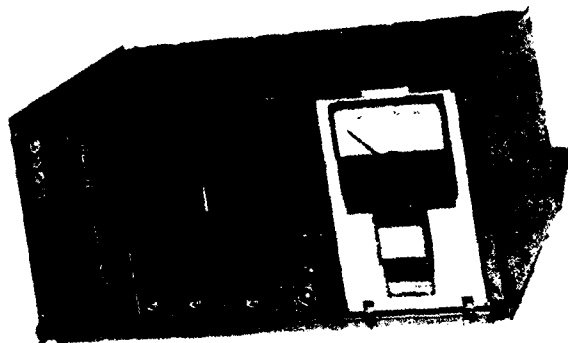


Fig. 9

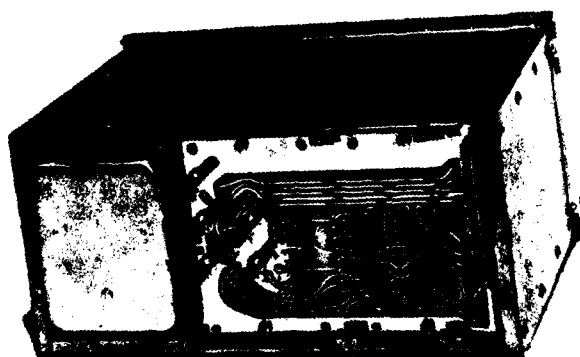


Fig. 10

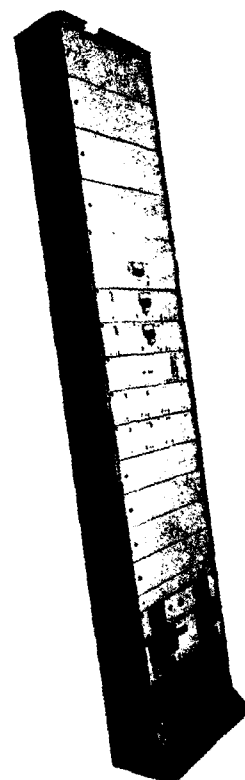


Fig. 11

TIME DETERMINATION FOR SPACECRAFT USERS
OF THE NAVSTAR GLOBAL POSITIONING SYSTEM (GPS)

Thomas J. Grenchik
NASA/Goddard Space Flight Center

Bertrand T. Fang
EG&G/Washington Analytical Services Center, Inc.

Summary

NASA is investigating the possibility that earth-orbiting artificial satellites could be users of GPS. The SEASAT spacecraft, to be launched into a 108 degree inclination, 800 km altitude earth orbit in 1978, was to have provisions for the reception of one way pseudo-range and pseudo range-rate measurements from the Phase I GPS satellites. Although the SEASAT/GPS experiment was cancelled because of time constraints, the present study still uses SEASAT as a specific example to investigate the effect of spacecraft clock offset on the GPS orbit determination capability, the optimum processing of measurements for the determination of clock offset, and the accuracy that can be obtained. The clock error enters the orbit determination process in the following two ways: as time tag error, and as signal propagation error. The problem of spacecraft as users of GPS is distinct from many other types of users in that the spacecraft state includes, as a minimum, three position and three velocity components plus the parameters defining the clock offset, and the comparatively predictable user vehicle dynamics provide smoothing for the determination of time offset.

The investigation assumes a minimum variance batch orbit determination process is used to solve for the orbital and clock parameters. To be realistic, in addition to the timing error, the GPS ephemeris error and the SEASAT dynamic modeling error are also taken into consideration. The results of the study indicate: the time tag error is of secondary importance, the optimum smoothing duration is of the order of one SEASAT orbit (≈ 100 minutes) and represents a compromise between the long duration desired for observability and the short duration desired to minimize dynamic errors, and if all errors considered are random in nature, the time offset and the SEASAT orbit may be determined to very high accuracy. However, modeling errors, although unknown, are systematic in nature. The correlation between the systematic errors and the clock offset limits the accuracy in time determination. Alternatively one may say the time offset accuracy imposes a requirement on such factors as our knowledge of the GPS ephemeris and the earth's gravitational field.

Introduction

The final presentation of calculated error can be clarified by digressing slightly. An unembellished review of basic concepts is appropriate at this point and is intended to lead to a rudimentary grasp of navigation error and the method of calculation. The discussion will cover in a simplified fashion, the following topics:

- The GPS measurements
- The concept of a mathematical model defining user spacecraft motion and its predictability.

- Orbit determination with the measurements and the model
- The relationship of error in user navigation to errors in the mathematical model and errors in the measurements
- The presentation of the results, the error in user position, velocity, and clock offset

GPS Measurements

A primary concept of this presentation is: Global Positioning System (GPS) navigation is performed by time measurements. If one were to err on the side of simplicity, GPS navigation in space could be cast as: "If you tell me where you are and what time it is, I will know where I am, and of course, I will know what time it is." A well-run train passes a station, the stationmaster bellows as I pass on the train, "It is 5:00 p.m. and this is Trenton." I set my watch and expect to be in Philadelphia at 5:40 p.m. The stationmaster is the GPS space vehicle and I am the user. Space travel is very predictable, the example of the train travel may not be, but it is meant only as an over-simplified example.

Two Body Model of Spacecraft Motion

In any field of endeavor there are a few basic principles, and there are thousands of niggling details concerning the basic principles. These are the details which complicate the subject, often beyond belief or comprehension. In describing the mathematical model of spacecraft motion, these details will be omitted. This simplified model will be the motion of two bodies in an isolated space. (Fig. 1)

A user, in some manner, has been placed in orbit about the earth. The earth is assumed to be spherically symmetric, the user is above the appreciable atmosphere, the sun and moon attract the user negligibly, and solar radiation has insignificant effect on the user trajectory. The only force exerted on the user is the force of attraction by the mass of the earth. The mass of the user is negligible compared to the mass of the earth. In this simplified situation it is possible to precisely define this user's motion with time.

It is convenient to construct a Cartesian coordinate system in which to uniquely describe the position and velocity of the user. (Fig. 2) One reference direction in the coordinate system is selected as the spin axis of the earth. The reference plane is the earth's equatorial plane, and a reference direction in this plane is made to point to a "fixed" pattern of stars on the celestial sphere. The third reference direction is chosen to form a right-handed, orthogonal coordinate system. The center of the coordinate system lies at the center of mass of the earth. Newton's laws are valid in this non-rotating, non-accelerating coordinate system, and there exists a

proportionality between the force of attraction acting on the user and the acceleration he experiences. The resulting set of three, second order differential equations of motion requires six initial conditions for solution; an initial position (3 components) and an initial velocity (3 components). The two body motion of the user about the earth is determined for all time by the solution of this set of equations.

NEWTON'S SECOND LAW APPLIED TO THE TWO BODY ORBIT

FORCE = (MASS) (ACCELERATION)

$$-k \frac{m_E m_S}{R^3} \vec{R} = m_S \frac{d^2 \vec{R}}{dt^2}$$

$$\frac{d^2 x}{dt^2} = \frac{-k m_E x}{(x^2 + y^2 + z^2)^{3/2}}$$

$$\frac{d^2 y}{dt^2} = \frac{-k m_E y}{(x^2 + y^2 + z^2)^{3/2}}$$

$$\frac{d^2 z}{dt^2} = \frac{-k m_E z}{(x^2 + y^2 + z^2)^{3/2}}$$

m_E = mass of the earth

m_S = mass of the user

k = constant of proportionality

$$R = \sqrt{x^2 + y^2 + z^2}$$

Orbit Determination

Orbit determination is the process of inferring the position, velocity, and clock offset of the user from measurements made of the user motion in the Newtonian coordinate system. The output of this process are a time or epoch, the three components of position at that time, the three components of velocity at that time, and user clock offset at that time. With the mathematical model of motion (the two body model of user motion), and the estimated position, velocity, and clock offset, the user can predict his trajectory for the present and future time. This result of knowing where you are and where you are going, is the process of navigation.

For GPS one of the measurements made of user motion can be described in the following way: (Fig. 3) The GPS space vehicle transmits to the user the GPS space vehicle time, position, and velocity. The user carries a clock, and if it is keeping time perfectly with the space vehicle clock, the user can measure the span of time required for the information to travel to him. The user, by scaling this delay time by the speed of light, computes the distance between the user and the space vehicle which sent the information. If the user clock and the space vehicle clock are offset, the user computes the distance between himself and the space vehicle, plus an additional distance related to the clock offset. In the calculation of distance, the GPS space vehicle position is known, since the space vehicle has already sent that information to the user.

IF SV CLOCK AND USER CLOCK ARE IN AGREEMENT:

$$(C) \text{MEASUREMENT} = \sqrt{(x_{SV} - x_{USER})^2 + (y_{SV} - y_{USER})^2 + (z_{SV} - z_{USER})^2}$$

$$\left(\frac{\text{Speed of Light}}{\text{Light}} \right) (\text{Delay Time}) = \text{Distance between space vehicle and user}$$

IF SV CLOCK AND USER CLOCK DIFFER BY A TIME OFFSET, Δt :

$$(C) \text{MEASUREMENT} = \sqrt{(x_{SV} - x_{USER})^2 + (y_{SV} - y_{USER})^2 + (z_{SV} - z_{USER})^2} + C \Delta t$$

$$\left(\frac{\text{Speed of Light}}{\text{Light}} \right) (\text{Delay Time}) = \text{Distance between space vehicle and user} + \left(\frac{\text{Speed of Light}}{\text{Light}} \right) \left(\frac{\text{User Clock Offset}}{\text{Offset}} \right)$$

x_{SV} = x coordinate of space vehicle

x_{USER} = x coordinate of user

\vdots

Δt = user clock offset

A measurement of the rate of change of time delay between the space vehicle and the user contains information about the space vehicle and user's velocities. This measurement is made in companion with the time delay measurement. Additional space vehicles supply non-redundant measurements, and when a sufficient number of measurements have been made, the user may begin his estimate of position, velocity, and clock offset.

Seven independent measurements from one or more GPS spacecraft enable one to solve for a set of preliminary estimates of the user orbital position, velocity and clock bias. The accuracy of such estimates is inadequate because of the inevitable measurement noise and the uncertainty in our knowledge of the GPS space vehicle orbits and the earth's gravitational field (the total mass of the earth and other geopotential coefficients describing the deviation of the earth from a homogeneous sphere). Therefore, precision orbits are usually determined by "curve-fitting" a user trajectory to a set of redundant measurements distributed over a chosen data span in accordance with the following procedure:

1. Calculate, from the preliminary estimate and the dynamic model of user motion, a trajectory for the user over the data span.
2. Calculate what the measurements would be, based upon the above trajectory and the model of the measurement process.
3. Relate the difference between the actual measurement and the calculated measurement to an adjustment in the preliminary estimate of user position, velocity, and clock offset.
4. Continue the process based on the current estimate until negligible adjustments are made to the estimate of user position, velocity, and clock offset.

The mathematics involved in this procedure will be described briefly in the following. Let $x(t_0)$, $y(t_0)$, $z(t_0)$, $\dot{x}(t_0)$, $\dot{y}(t_0)$, $\dot{z}(t_0)$ represent the preliminary estimate. " t_0 " is some reference time, which will be chosen as initial time for the convenience of the following discussion. The user dynamic equations of motion, similar to but more complete than the two body equations, are then integrated, subject to the preliminary estimate as initial conditions, to obtain a user trajectory

$$x_C(t) = f(x(t_0), y(t_0), z(t_0), \dot{x}(t_0), \dot{y}(t_0), \dot{z}(t_0); \alpha)$$

with similar expressions for the other position and velocity components. The computed trajectory is not the true user trajectory because the initial estimate $x(t_0)$...etc. are in error and because of uncertainties in our knowledge of the dynamic parameter α , which stands for geopotential coefficients, atmospheric drag and etc. To first order one may relate the trajectory error to the error sources as

$$\Delta x_C(t_0) = \frac{\partial f}{\partial x(t_0)} \Delta x(t_0) + \frac{\partial f}{\partial \dot{x}(t_0)} \Delta \dot{x}(t_0) + \dots + \frac{\partial f}{\partial \alpha} \Delta \alpha$$

From the computed trajectory, the range from the user to a GPS spacecraft may be calculated as

$$R_C(t) = \sqrt{(x_{SVC} - x_C(t))^2 + (y_{SVC} - y_C(t))^2 + (z_{SVC} - z_C(t))^2}$$

which is equivalent to a signal transit time of $T_C(t)$

$$T_C(t) = \frac{R_C(t)}{C}$$
 On the other hand, the measured signal transit time may be written as

$$\begin{aligned} T(t) &= \frac{1}{C} \sqrt{(x_{SV} - x(t+\Delta t))^2 + (y_{SV} - y(t+\Delta t))^2 + (z_{SV} - z(t+\Delta t))^2} \\ &\quad + \Delta t + n(t) \\ &\approx T_C(t) + \frac{1}{C} \left\{ \frac{\partial R_C(t)}{\partial x_C(t)} \Delta x(t) + \frac{\partial R_C(t)}{\partial x_{SVC}} \Delta x_{SVC} + \dots \right. \\ &\quad \left. + \left\{ \frac{1}{C} \frac{\partial R_C(t)}{\partial x_C(t)} \dot{x}_C(t) + \dots + 1 \right\} \Delta t + n(t) \right\} \end{aligned}$$

where C is velocity of light and $n(t)$ is random measurement noise. By making use of the above equations, the difference between the actual and the computed measurement may be written as

$$\begin{aligned} T(t) - T_C(t) &= \frac{1}{C} \left\{ \frac{\partial R_C(t)}{\partial x_C(t)} \left(\frac{\partial f}{\partial x(t_0)} \Delta x(t_0) + \frac{\partial f}{\partial \dot{x}(t_0)} \Delta \dot{x}(t_0) \right. \right. \\ &\quad \left. \left. + \dots + \frac{\partial f}{\partial \alpha} \Delta \alpha \right) \right. \\ &\quad \left. + \dots + \frac{\partial R_C(t)}{\partial x_{SVC}} \Delta x_{SVC} + \dots \right\} \\ &\quad + \left\{ \frac{\partial R_C(t)}{\partial x_C(t)} \frac{\dot{x}(t)}{C} + \dots + 1 \right\} \Delta t + n(t) \quad (1) \end{aligned}$$

For the data span considered, there are many other equations similar to this, representing different measurements made at different times and from different GPS spacecraft. The left-hand side of these equations are known quantities which serve as feedback error signals. On the right-hand sides are the Δ 's which represent two kinds of quantities. The user orbit estimate errors such as $\Delta \dot{x}(t_0)$, $\Delta x(t_0)$ and the clock offset Δt are quantities to be inferred from these feedback signals. The random noise $n(t)$, the dynamic parameter uncertainty $\Delta \alpha$, and the GPS spacecraft ephemeris error such as Δx_{SV} are unknown error sources. Although the latter quantities appear in these equations not any differently from the former, the limited information content and the computational constraints preclude the estimation of every unknown parameter appearing in these equations. Before giving an expression for the feedback gain, the following general observations may be made:

1. The error sources not estimated will cause some errors in the estimated parameters. Random noises tend to be smoothed out because there will generally be many more measurements than the parameters to be estimated. At the current state of art, systematic errors such as GPS ephemeris error, geopotential error, and atmospheric drag error at low altitudes, are more important. The latter two are dynamic modeling errors which generally cause errors that amplify with elapsed time.

2. The observable "feedback error signal" must be sensitive to variations in a parameter " Δ " at a detectable level and in a unique way, for the inference of that parameter to be possible. This implies:

a. Ambiguity exists in estimated parameters if they have similar effects on the observed "error signal."

b. In general, $\frac{\partial f}{\partial x(t_0)} \left(\frac{\Delta \partial x(t_0)}{\partial x(t_0)} \right)$ and therefore the sensitivity of the observed error signal to orbit corrections $\Delta x(t_0)$, increases with elapsed time $(t-t_0)$. From this consideration a long data span is desirable. This also applies to the estimation of spacecraft clock drift-rates. On the other hand, as discussed in "1" above, a long data span is susceptible to unmodelled dynamic parameter errors. Thus a compromise has to be made.

c. The "unity" in the coefficient of Δt represents the transit time error while the rest of the terms represent the time-tag error. Obviously the latter is negligible in comparison with the former as far as range measurements are concerned. Transit time error is the error in the measured delay time. Time tag error is the error in relating the measurement time to epoch.

Let us group together all the measurements in the following matrix form which is a generalization and rearrangement of Equation (1):

$$[y-y_C] = [A] \begin{bmatrix} \Delta x(t_0) \\ \Delta y \\ \Delta z \end{bmatrix} + [B] [\Delta \delta] + [N] \quad (2)$$

where

Δy = parameters to be estimated such as the user spacecraft clock error.

$\Delta\delta$ = parameters not estimated such as the GPS space vehicle ephemerides errors and some geopotential coefficient errors.

N = zero-mean random measurement and dynamic noise.

$y-y_c$ = feedback error signal, the difference between the actual and computed measurements.

$[A], [B]$ = known partial derivatives representing the sensitivities of the error signal with respect to the parameters to be estimated and to those error sources not estimated.

The corrections to the preliminary estimate are obtained from the error signal as

$$\begin{bmatrix} \Delta x^*(t_o) \\ \Delta y^* \end{bmatrix} = [K] [y-y_c] \quad (3)$$

where the feedback gain is usually chosen to be

$$[K] = ([A]^T [W] [A])^{-1} [A]^T [W] \quad (4)$$

$[W]$ = chosen positive-definite weighting matrix.

As stated before this procedure is repeated until convergence occurs; i.e., $\begin{bmatrix} \Delta x^*(t_o) \\ \Delta y^* \end{bmatrix} \approx \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. The remaining orbit determination or estimation error may be obtained from Equations (2) and (3) as

$$\begin{bmatrix} \Delta x(t_o) \\ \Delta y \end{bmatrix} = -[K] ([B][\Delta\delta] + [N])$$

The unmodelled parameter error $[\Delta\delta]$ and the random noise $[N]$ in this expression are of course unknown. But once their statistical properties such as means, variances and correlations are specified based on experience and deduction, the corresponding statistical properties of the estimate may be obtained immediately. The variances are measures of the accuracy of the estimate such as the confidence one has in the spacecraft clock offset estimate. The correlation coefficients, if reaching close to unity, would signify that some of the parameters may not be estimated without ambiguity.

The use of the gain matrix given in Equation (4) results in the so-called weighted least-squares estimate because the estimate minimizes the weighted squares of the error signal; i.e., $[y-y_c]^T [W] [y-y_c]$. Frequently the weighting matrix is chosen as the inverse of the covariance matrix of the measurement error. In the ideal, but unrealistic situation in which all error sources are random in nature, this choice minimizes the variance of the estimate, and one has the so-called un-biased minimum variance estimate.

User Position, Velocity, and Clock Error

To illustrate the effect of clock errors and the accuracy with which the USER spacecraft time and orbit may be determined, a low-earth-orbit spacecraft (SEASAT) as tracked by six Phase I GPS space vehicles is considered. The orbital characteristics of the user and the space vehicles are given in Table 1. Error sources corresponding to the present state of art are described in Table 2. For this tracking configuration and error model, the computer program

ORAN is used to simulate a batch minimum-variance, weighted least-squares orbit determination process and compute the corresponding errors and accuracies. The results are summarized in the following paragraphs.

1. Orbit Determination Accuracy. For all but very short tracking spans, random measurement errors are negligible in comparison with the unmodelled parameter errors. The tracking rate is also not very important. For short tracking-span orbit determination geometric errors (primarily GPS ephemeris errors) dominate. Dynamic errors (primarily geopotential uncertainty) tend to grow with time and become important for long tracking span orbit determination. The tracking time span for optimum orbit determination is around one user orbital period (≈ 100 minutes). The resulting user orbit uncertainty has standard deviations of approximately 8 meters in position and 1 cm/sec in velocity. About 3/4 of these are the contribution of geometric error sources.

2. Effect of Time Tag Errors. It may be seen from Eq. (1) that the effect of time tag errors is proportional to the spacecraft velocity and acceleration. Our result indicates that a time-tag error of 100 microseconds contributes approximately 1/2 meter orbit position error. Since time-tag accuracy is expected to be much better than 100 microseconds one may say the effect of time-tag error is negligible. Because its effect is small, it is also not possible to estimate time-tag error with accuracy.

3. Clock Bias Estimation. USER spacecraft measures its range from a GPS space vehicle by keeping track of the signal propagation time. A 10 nanosecond clock bias is equivalent to 3 meter range bias.

Our results indicate the following with regard to estimation of clock bias. In the absence of unmodelled dynamic parameter errors clock biases may be determined to the nanosecond level. There is, however, a high correlation between the clock bias and the uncertainty in the gravitational parameter GM, i.e., the product of the universal gravitational constant and the total mass of the earth. It is, therefore, not possible to determine clock bias to better than 25 nanosecond accuracy in the presence of a gravitational error of one part per million.

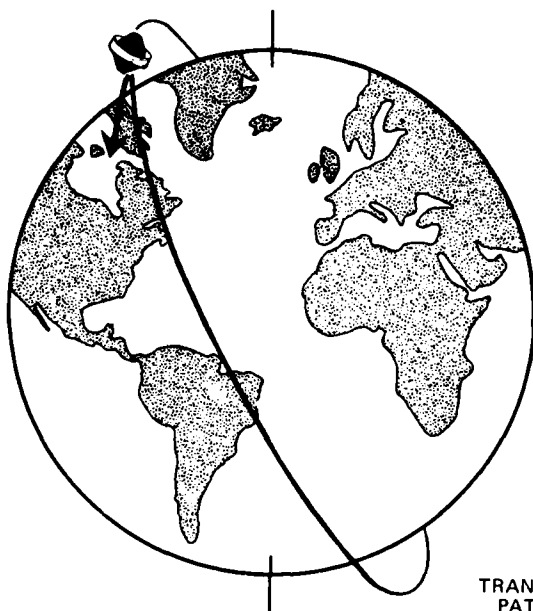


Fig. 1 MOTION OF USER SPACECRAFT ABOUT EARTH

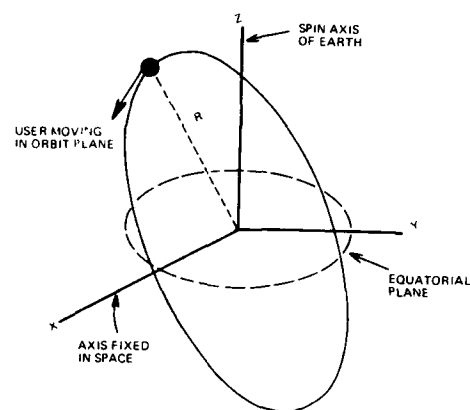


Fig. 2 NEWTONIAN COORDINATE SYSTEM:
NON-ROTATING, NON-ACCELERATING

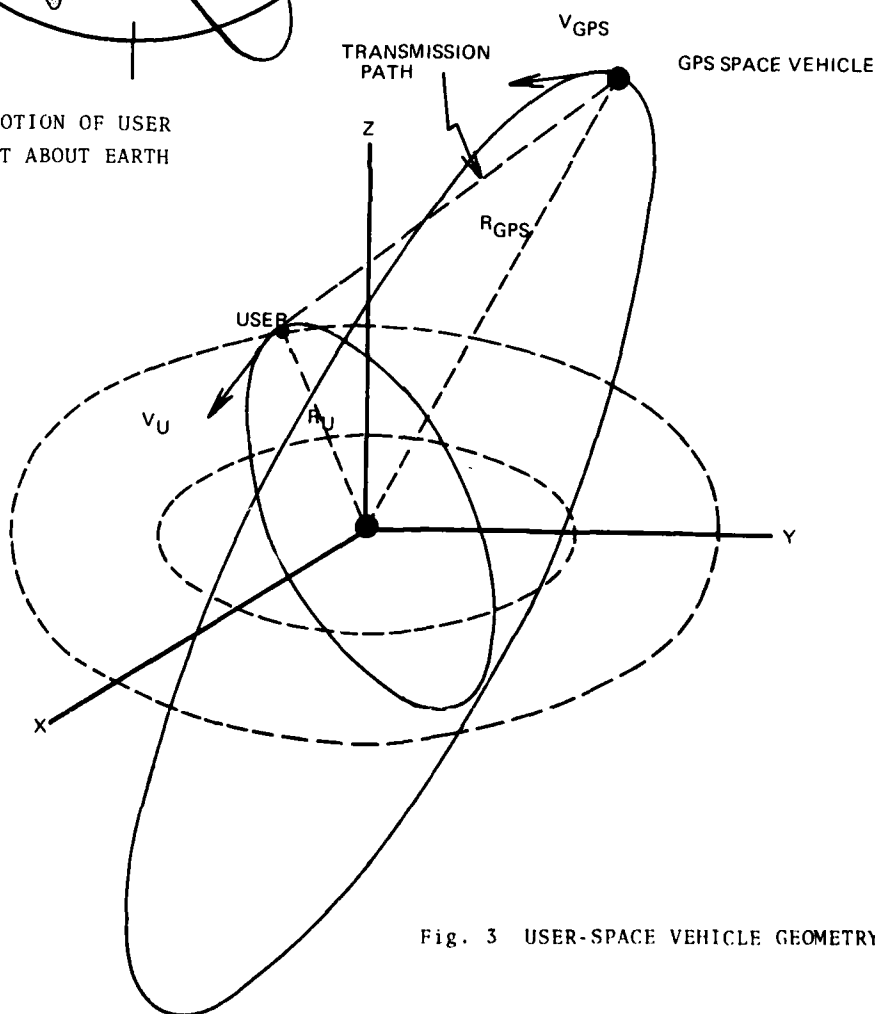


Fig. 3 USER-SPACE VEHICLE GEOMETRY

TABLE 1. ORBITAL CHARACTERISTICS OF PHASE I
GPS SPACECRAFT AND USER (SEASAT)

SPACECRAFT	INCLINATION	PERIOD	NODE	MEAN ANOMALY
GPS #1	63°	12 Hrs.	-120°	-23.2°
#2			-120°	16.8°
#3			-120°	56.8°
#4			120°	16.8°
#5			120°	56.8°
#6			120°	96.8°
USER (Circular Orbit at 800 Km Altitude)	108°	100.5 Min.	-75°	180°

TABLE 2. ERROR MODEL

Standard Deviation of Tracking System Error	
Range Noise	1 m
Range-Rate Noise	0.15mm/s
GPS Spacecraft Ephemeris Error	30 m (Position) 0.5 cm/s (Velocity)
Standard Deviation of Dynamic Parameter Uncertainty	
Atmospheric Drag	25%
Earth GM	1 part per million
Geopotential Harmonics	50% of the difference between the two gravity models, MD1 and GM5

TRANSCONTINENTAL AND INTERCONTINENTAL PORTABLE CLOCK TIME COMPARISON

H. Hellwig, D. W. Allan, S. R. Stein and K. A. Prichard
Frequency & Time Standards Section
National Bureau of Standards
Boulder, Colorado 80302

NBS has developed and used a small portable rubidium clock, which permits time comparisons with other laboratories at much less cost than possible with the traditionally used cesium clocks [1, 2]. The use of this clock over the past year has shown that time generating clock ensembles perform with stabilities of parts in 10^{14} for time periods of months. National and international precision time comparisons are done today routinely via Loran-C and portable atomic clocks. Satellite links are being explored today and may become the main link of the future. Nevertheless, portable clocks will be needed to calibrate satellite links and to study variations in such transmissions.

The stability of the rubidium clock and its sensitivity against environmental effects is almost an order of magnitude inferior to that of cesium clocks. However, in successful round trips, using time closure between departure and arrival at the originating laboratory, very acceptable time comparison precision can be achieved. More frequent portable clock comparisons not only lead to better confidence in the time comparison and more information on the Loran-C link but also to an increased ability to compare frequencies between time generating laboratories even when separated by large distances. With a time accuracy of 100 ns and clock trips spaced three months apart, fractional frequency measurement precision of 10^{-14} is actually achieved by our series of time comparisons as reported below.

An important difference must be noted between these rubidium portable clock trips and trips possible using available cesium clocks. Because of the inferior stability of the rubidium clocks and the greater environmental sensitivity, both almost a factor of 10 inferior as compared to cesium, clock trips must be done as rapidly as possible. If time is to be carried to a remote station which does not have a high precision timing system on site, an immediate return of the clock and a quick trip are mandatory for good accuracy. However, if the laboratory to be measured or to be compared with is itself a high precision time-keeping laboratory, a quick round-trip becomes much

less important. Upon arrival of the portable clock, the clock is measured against that laboratory's time signals, monitored during its stay there, and measured immediately before leaving for its return trip. The portable clock could remain at such a laboratory for times of the order of one week, if the frequencies are a priori known to within 10^{-13} . Despite the advantages in weight, size, ease of handling, etc., it is obvious that the rubidium clock would be much inferior to a cesium clock for a long-term clock transport involving a series of remote sites, each having not very well-known time or frequency signals [3,4].

We executed numerous trips from the National Bureau of Standards (NBS) in Boulder, Colorado to the U.S. Naval Observatory (USNO) in Washington, D.C. and several comparisons with the Bureau International de l'Heure (BIH) in Paris, France, during the past year. These trips served to bring about a comparison of the respective time scales with previously unachieved precision. They also served to check the capability of the portable rubidium clock. Support for the capability of rubidium portable clocks, as quoted above, is directly given by the data points in figures 1 and 2. In January 1977 and April 1977, we carried two rubidium clocks simultaneously on the same trip to the BIH. In fact, it was possible for one person to carry two portable rubidium clocks with no major difficulty to the BIH in Paris.

In figures 1 and 2, we give the measurement results of a series of portable clock trips between NBS and USNO and NBS and BIH respectively. Plotted in figures 1 and 2 are the new 1977 rates; i.e., the actual rate changes were used to modify all data prior to 1 January 1977. All data on rubidium clocks are based on NBS records, all data on cesium clocks are based on USNO records.

When using portable clock data to compare the time and frequency of two time scales remote to each other, it is profitable to consider the general character of the instabilities involved which may limit the comparison. In figure 3 we illustrate graphically an estimate of the current instabilities in the USNO and NBS time scales, including the effect of the linear frequency drift apparent in figure 1 on $\sigma_y(\tau)$ (the two-sample deviation). We assume that other time scales have

instabilities of the same order. Also plotted in figure 3 is the portable clock comparison measurement noise achievable from repeated portable clock trips; we use an rms time error for each trip of 90 ns, which was obtained from figure 1. Therefore, the noise-model of figure 3 is applicable only to the USNO/NBS comparison of time scales.

From figure 3 it appears that for the most part the predominant "signal" process between two time scales is frequency drift; whereas the predominant measurement limitation is the accuracy achievable with repeated portable clock trips. This limitation may be well modeled by a white phase (or time) noise process, hence a good estimate of the comparison of two time scales via repeated portable clock measurements is a least squares quadratic fit to the time difference measurements. The curves in figures 1 and 2 are a result of such a fit in the comparisons between UTC - UTC(NBS) and UTC(USNO) - UTC(NBS). If one analyzes the data of figure 1 as departures from the fit due to random instabilities other than the assumed linear frequency drift between the time scales involved, then one obtains the very impressive stability $\sigma_y(\tau \sim 3 \text{ months})$ between the scales of 1.5×10^{-14} . The error and stability for the BIH/NBS link is not as good as can be seen from figure 2.

The time differences as given by Loran-C measurements are plotted in figure 4. The error and stability associated with the Loran-C data are significantly worse than those achievable by repeated portable clock trips. In fact, one observes significant departures of the Loran-C measurements from the least-squares quadratic fit over the course of a year of almost 1 μs for the USNO/NBS data and 1.5 μs for the BIH/NBS data. Also, the data strongly suggest seasonal variations in both Loran-C links.

In summary, we submit that we have shown that time scales can be compared in frequency to parts in 10^{14} via repeated portable clock trips and that the corresponding capability of the Loran-C links is about 1×10^{-13} . The data also indicate, that time scales show drifts of the order of 1×10^{-13} per year. It appears futile to attempt scientifically sound explanations of such drifts. However their presence illustrates the importance of a steering of TAI (and UTC) using data from primary standards. This steering is being implemented [5] by the BIH. Our data show that, at present, Loran-C links would not be sufficient to relate the data needed for the steering from the contributing primary standards laboratories to the BIH and to compare primary standards in these

laboratories. Portable clocks must therefore be used for these purposes until it is demonstrated that a substitute system such as a satellite based system has equal or better performance. We hope that both small rubidium and cesium portable clocks will become commercially available so that laboratories involved in time-keeping and coordination can take advantage of the available time comparison accuracy.

ACKNOWLEDGMENT

The authors of this paper who were involved in the actual data interpretation, data reduction and clock development would like to express their sincere appreciation for the cooperation of many people and laboratories; they made our clock trips successful, indeed possible. Their contributions include time measuring and monitoring of our clocks at the USNO and the BIH, the help on transport to and from the airport, and in many cases the actual carrying of clocks between BIH, NBS and USNO, often including special arrangements to accommodate the clock trips. The authors are very gratified that a true international cooperating community of scientists and engineers in this field exists. Without this cooperation, our results would not have been possible.

Dr. G. M. R. Winkler and Dr. G. Hall of the USNO kindly supplied us with the data on cesium portable clocks and extended special help to us on many occasions.

REFERENCES

- [1] "Sub-Microsecond Time Transport with a Rubidium Portable Clock," Helmut Hellwig and A.E. Wainwright, Proc. 29th Annual Symposium on Frequency Control, Atlantic City, NJ, May 1975, pp. 384-386.
- [2] "A Portable Rubidium Clock for Precision Time Transfer," Helmut Hellwig and A.E. Wainwright, Proc. 7th Precise Time and Time Interval (PTTI) Planning Meeting, Washington, DC, December 1975, pp. 143-159.
- [3] "Around-the-World Atomic Clocks: Predicted Relativistic Time Gains," "Around-the-World Atomic Clocks: Observed Relativistic Time Gains," J.C. Hafele and Richard E. Keating, Science, Vol. 177, July 1972, pp. 166-170.
- [4] "Recent Experiments with Flying Atomic Clocks, Loran-C, Omega, and VLF for Clock Synchronization, Gernot M.R. Winkler, Report to XVII General Assembly, Warsaw 1972, of the International Union of Radio Science.
- [5] 8th Session of the Consultative Committee for the Definition of the Second, Sevres, France, 26-27 April 1977, Recommendation S1
- [6] NBS Time and Frequency Services Bulletin, No. 230, p. 7, January 1977.
- [7] USNO Time Service Announcement Series 14, No. 11, 27 October 1976.
- [8] Bureau International de l'Heure Circular D124, 3 March 1977.

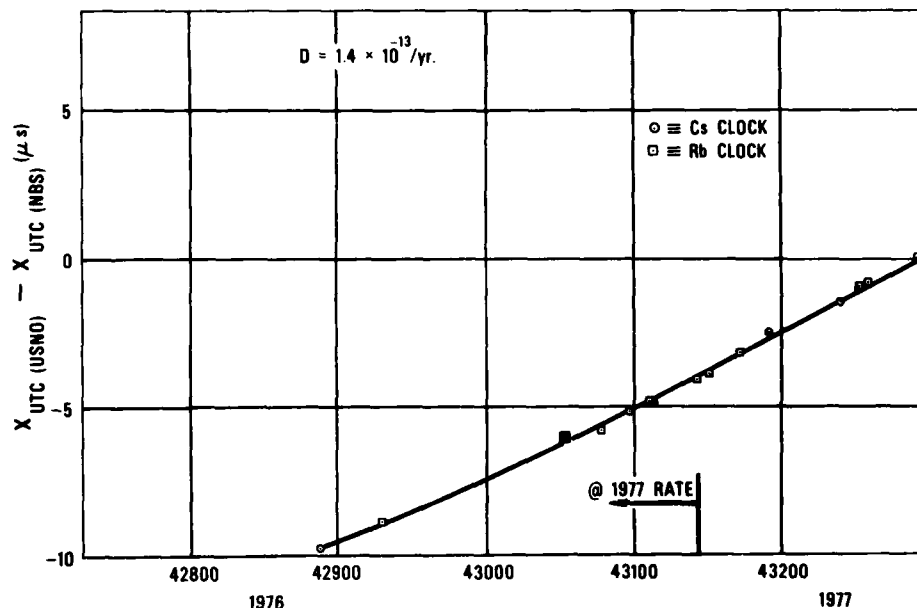


FIGURE 1 Portable clock comparisons between USNO and NBS. The curve, $X(n)$, is a least squares, quadratic fit: $X(n) = (16.27 \text{ ps/d}^2)n^2 + (25.60 \text{ ns/d})n - 4.256 \text{ } \mu\text{s}$, where $n = 1977$ day count. D is the fractional frequency drift per year. All available rubidium and cesium data are plotted. The rubidium measurement around MJD 43080 is due to simultaneous transport of two independent clocks, yielding agreement to within 50 ns. The data are "normalized" to the 1977 time scale rates; i.e. the 1976 data were adjusted by 72 ns/day for UTC(NBS) [6] and 60 ns/day for UTC(USNO) [7]. Horizontal axis = Modified Julian Day (MJD)

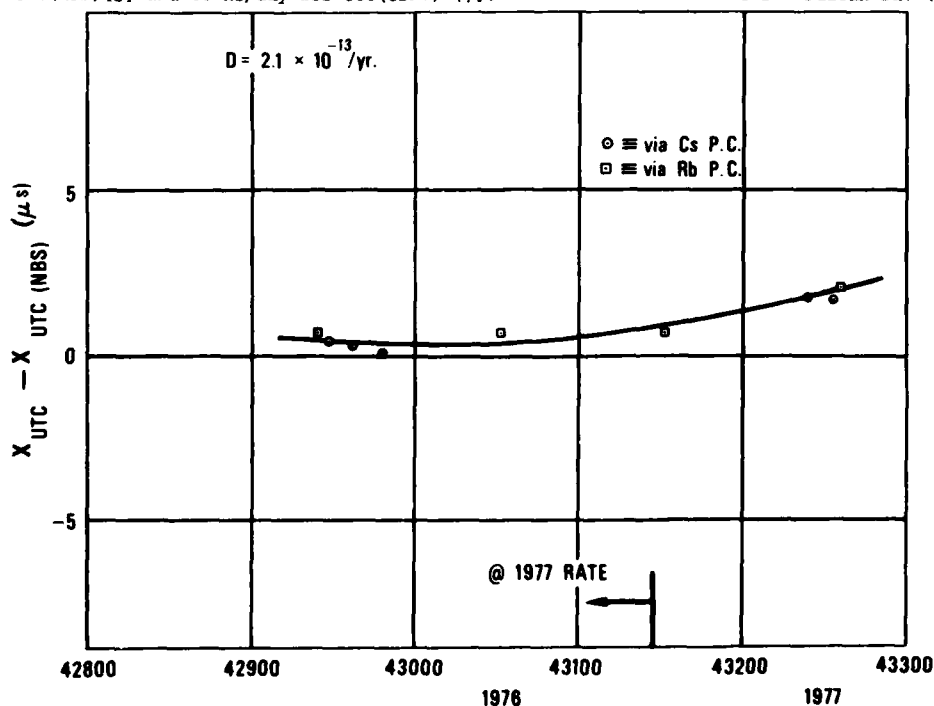


FIGURE 2 Portable clock comparisons between BIN and NBS. The curve, $X(n)$, is a least squares, quadratic fit: $X(n) = (22.81 \text{ ps/d}^2)n^2 + (6.64 \text{ ns/d})n + 0.837 \text{ } \mu\text{s}$ where $n = 1977$ day count. D is the fractional frequency drift per year. All available rubidium and cesium data are plotted. The rubidium measurements around MJD 43150 and MJD 43260 are due to simultaneous transport of two independent clocks each, yielding agreement to within 90 ns for each trip. The data are "normalized" to the 1977 time scale rates; i.e. the 1976 data were adjusted by 72 ns/day for UTC(NBS) [6] and 86.4 ns/day for UTC [6,7]. Horizontal axis = Modified Julian Day (MJD).

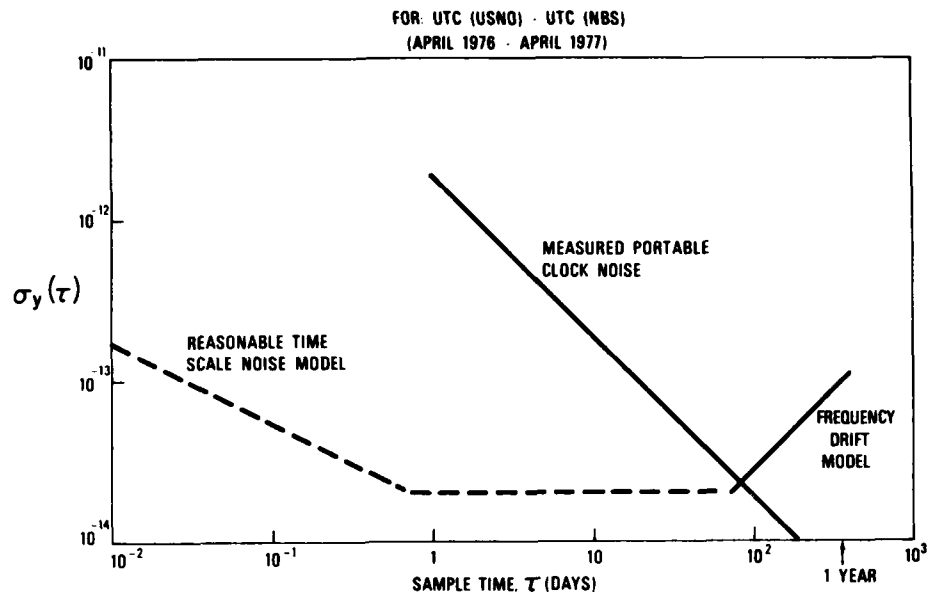


FIGURE 3 Noise performance of time scales and portable clock comparisons. The time scale noise model is derived from internal measurements at NBS and believed to be typical for today's high performance time scales. The portable clock noise and the time scale rate drift are calculated from the data of figure 1; therefore they only apply to the USNO/NBS comparison.

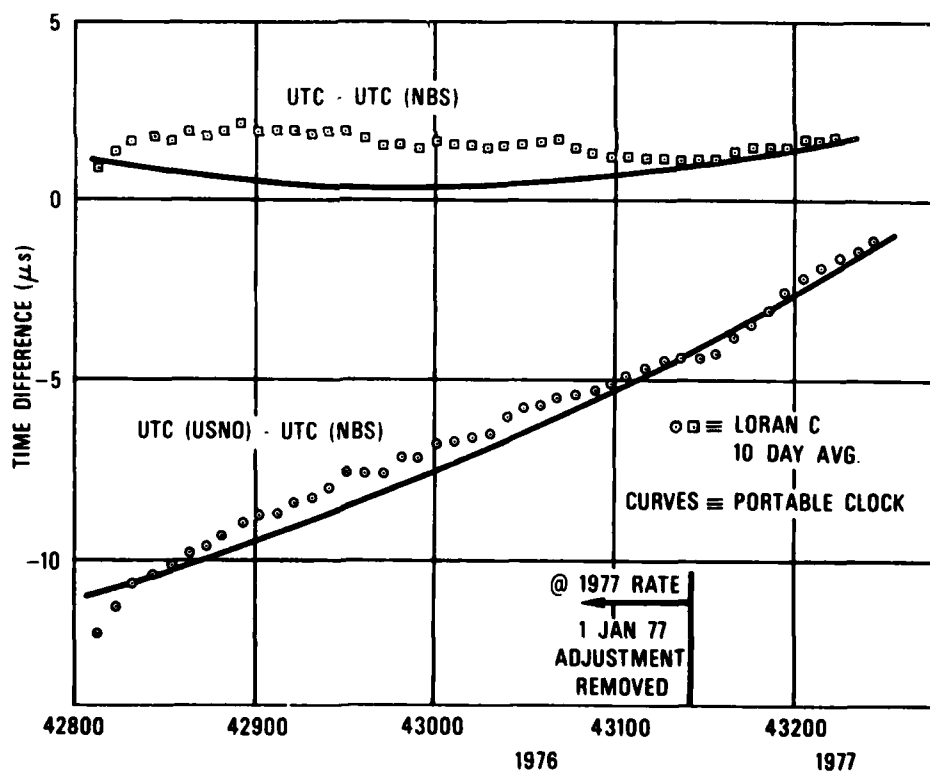


FIGURE 4 Comparison of Loran C data and portable clock data for the links USNO/NBS and BIH/NBS. The curves are identical to those in figure 1 and figure 2. The Loran C data are taken from the BIH Circular D. As in figure 1 and figure 2, the plotted rates are those of 1977; i.e., the 1976 data are adjusted. In addition, the 1976 data are also modified to remove a Loran C time-step adjustment by the BIH on 1 January 77 ($-0.4 \mu\text{s}$ for UTC(USNO) and $-0.2 \mu\text{s}$ for UTC(NBS) [8]). Horizontal axis = Modified Julian Day (MJD).

DISTANT COMPARISON OF STABLE FREQUENCY STANDARDS BY MEANS OF THE TRANSMISSION OF A BEAT NOTE BETWEEN THE CARRIER OF A TV BROADCAST SIGNAL AND A FREQUENCY SYNTHESIZED FROM THE FREQUENCY STANDARDS.

A. GABRY, G. FAUCHERON, B. DUBOIS

CENTRE NATIONAL D'ETUDES DES TELECOMMUNICATIONS

Issy-les-Moulineaux,

FRANCE

AND

P. PETIT

LABORATOIRE DE L'HORLOGE ATOMIQUE - UNIVERSITE PARIS-SUD

Orsay, France

Summary

In his method of distant comparison of stable atomic frequency standards, a low frequency beat note is produced between the carrier of a TV broadcast signal and a signal which is frequency synthesized from each frequency standard, in laboratories A and B. The audiofrequency signal obtained in laboratory A is transmitted to laboratory B via a telephone line. It is phase compared to the other audiofrequency signal which is locally produced in laboratory B. A record of the variation of the relative phases of the two frequency standards is then easily achieved. The frequency of the stable sources being converted up to the TV signal frequency (close to 185 MHz in the example given), one can easily see that the resolution of this method of phase comparison must be excellent. The allowed resolution of frequency comparisons is then equal to $4 \times 10^{-11} \times 1/\tau$ for $1 \text{ s} < \tau < 300 \text{ s}$ and 6×10^{-4} for $\tau = 1 \text{ hour}$.

Introduction

In order to compare different clocks, a single TV pulse can be used¹. This yields an accuracy of roughly 10 ns i.e. a relative precision of 10^6 for a measurement time of 1 s, or 10^{-11} for 1000 s or 10^{-13} for one day. This points out the fact that only long term measurements are significant with this method. High precision short term time and frequency comparisons requires at least one ns accuracy. This can be achieved using laser pulses², when both laboratories are directly visible from each other.

In this paper we describe a new and cheap method using a TV carrier together with a simple telephone line. It allows a precise comparison of the short term frequency stability of distant frequency standards. The experiment has been performed in France between Orsay and Issy-les-Moulineaux near Paris (Fig 1) using a 185 MHz TV Carrier. Although the first laboratory (Orsay) is not directly visible from the emitting antenna (Eiffel tower), the reception level is good enough.

Experimental Set-Up

A requirement of 1 ns precision is equivalent to a carrier phase precision of 0.4π . This can be easily obtained using only low frequency measurement, with the set up shown on figure 2.

The two frequency standards to be compared are both hydrogen-masers. The same nominal low frequency is synthesized in each laboratory, as shown on

figure 3, from the TV carrier and H-maser output signal frequencies. The telephone line is used to compare the two phases which should be equal if the frequency standards were identical.

Neglecting the electric length variations of the wire, the possible phase changes might come from: the TV set, the frequency synthesizer, the carrier reception and fluctuations of the propagation delay. Effect of the TV set.

The TV set is used as a mixer of the TV-video carrier and of the synthesized signal. Since it is a broadband device and owing to the low required bandwidth (a few 100 Hz corresponding to the low frequency signal), linear distortions can be neglected. It has been checked with respect of both tuning and reception level, that a good enough phase stability is achieved.

Synthesizers.

The synthesizers are very simple and then required phase stability is easy to achieve. With commercially available synthesizers, the phase of the low frequency signal depends upon the amplitude adjustment of the synthesized signal (specially at 176254.39 kHz) but this is a negligible effect since this amplitude is a constant.

Reception at Orsay.

According to the experiment, there is good evidence that the phase uncertainty of the received TV carrier is less than 0.04π (0.1 ns) in spite of the emitter antenna is not directly visible from there.

Propagation delay.

The two laboratories (Orsay-Faculté des sciences and Issy-les-Moulineaux-CNET) are located in the same direction with respect of the single TV transmitter (Paris-Eiffel tower). The distance between the laboratories is 16 km. The air refraction index may be evaluated from the following formula³

$$(n-1)10^6 = \frac{77.6}{T} \left(p + \frac{4810}{T} e \right)$$

where

- p (mB) is the atmospheric pressure
- e (mB) is the fractional pressure of water vapor
- T (K) is the thermodynamic temperature

Using the table shown on figure 4, this can be translated in terms of propagation delay versus humidity, assuming a normal atmospheric pressure as describe on figure 5. In the condition of stable weather, it is clear that the variations of the propagation delay can be neglected specially at low temperature. It should be noticed that the considered fluctuations of the propagation delay are a common limitation of all methods using such a propagation. It is obvious that the precision can be improved by applying the relevant corrections.

Experimental results

The dots on the figure 6 show that this method is able to achieve a precision in time measurement of the order of 40 ps for sampling time between one to several hundreds seconds. The curve bends for a delay greater than 300 seconds giving a less good precision. There the curve shows a -1/2 slope, instead of -1. It is not presently possible to decide if this loss of precision is rather a result of the measurement method than an instability in the masers themselves at the moment of the experiment. For the purpose of illustration figure 7 is a phase record between two distant hydrogen masers.

The reported then allows a resolution of frequency comparison which equals $4 \times 10^{-11} \times 1/\tau$ for $1s < \tau < 300s$ and 6×10^{-14} for $\tau = 1$ hour.

The method has been used to characterize the frequency stability of a stable cesium beam frequency standard in Issy-les-Moulineaux, by comparison with a hydrogen maser located 16 km apart in Orsay. We obtain $\sigma(\tau = 12 \text{ hours}) = 1.1 \times 10^{-13}$.

Alternatively, these results show that convenient and precise measurements of the frequency of a hydrogen maser are possible, relative to a distant cesium beam frequency standard, and therefore to the best realizations of the second, through TAI. Investigations and checks of methods intended for the improvement of the accuracy of the hydrogen maser can then be easily implemented, even in the absence of a nearly primary cesium beam frequency standard.

Conclusion

It has been experimentally proved that distant comparisons of frequency standards, through the carrier of a TV signal can be made with a precision which is equivalent to a short term resolution of 40 ps in time measurement. The possibility of long term frequency comparison can be contemplated in areas where TV stations broadcast 24 hours a day.

References

- 1 - Parcelier P. (LPTF - Observatoire de Paris) "Comparaison d'horloges atomiques par reception de signaux de television" Bulletin d'Information du Bureau National de Métrologie (FRANCE) Janvier 1976, numéro 23

- 2 - Besson J. et Parcelier P. "Synchronisation dans le domaine de la nanoseconde d'horloges éloignées". Proceedings of the 1 X th International Congress of Chronometry Stuttgart 16-20/9/1974

3. CCIR - Avis 453
XIII Assemblée plénière Genève 1974 Vol. 5 p. 65.

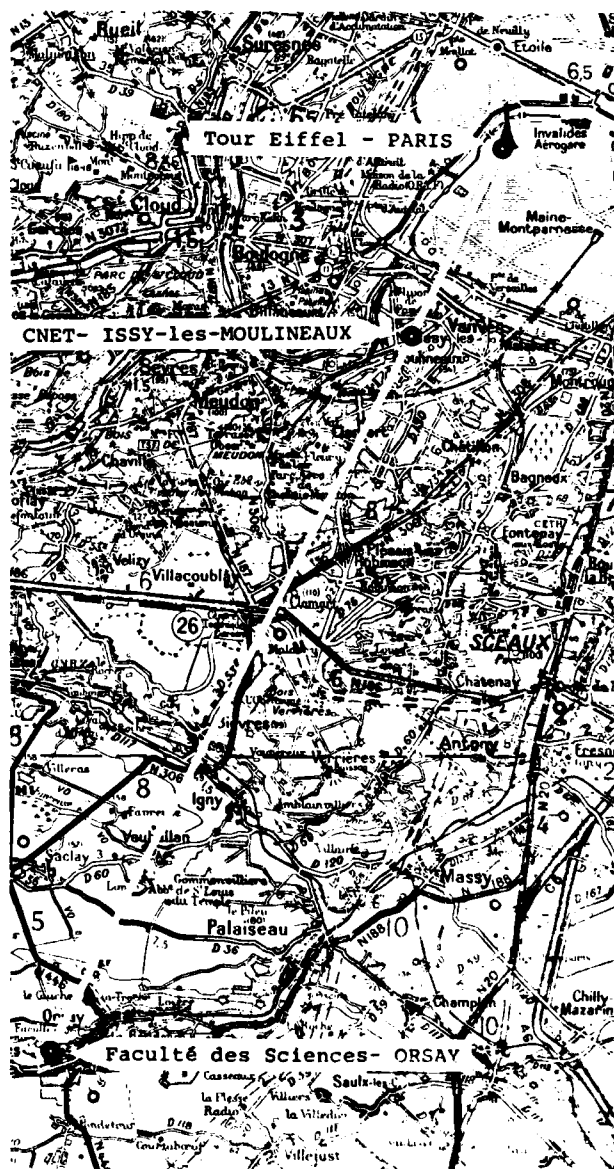


Figure 1 - Geographic position of laboratories and TV stations

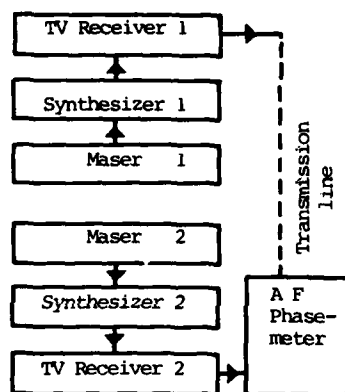


Figure 2 - Diagram of the measurement method

Temperature (degree c)	water vapor pressure (mB)
40	73,7
38	66,2
36	59,4
34	53,2
32	47,5
30	42,43
28	37,78
26	33,65
24	29,82
22	26,40
20	23,37
18	20,61
16	18,16
14	15,98
12	14,03
10	12,28
8	10,73
6	9,35
4	8,13
2	7,05
0	6,10

Figure 4 - Table of saturated water vapor pressure versus temperature

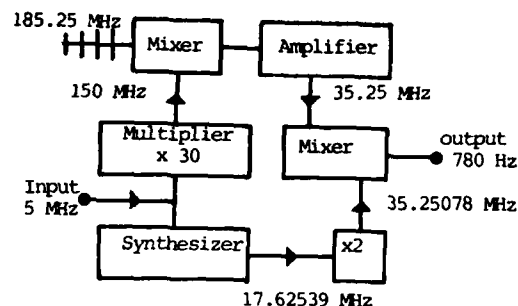


Figure 3 - Diagram of the beat-frequency generator

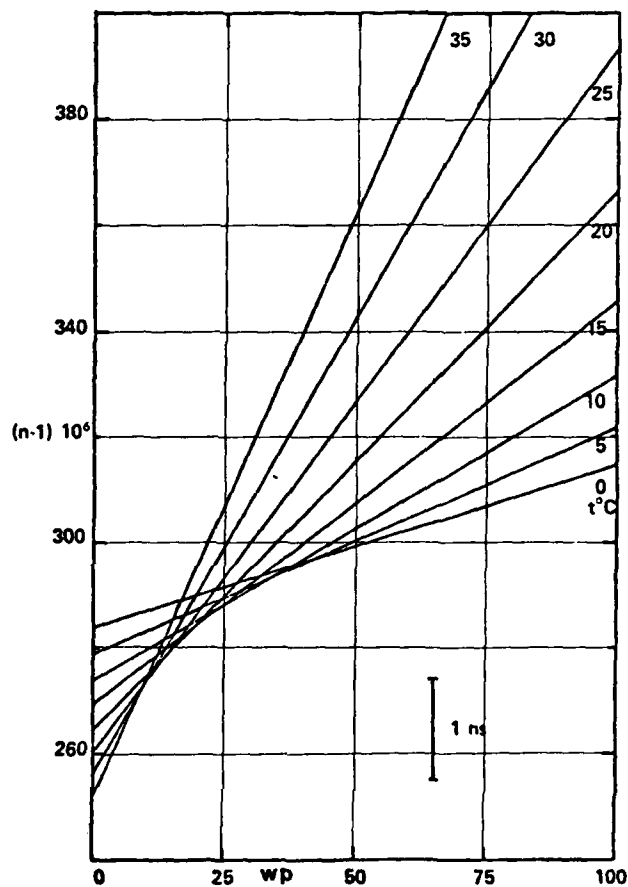


Figure 5 - Curves giving the air refraction index (n) versus the fractional pressure of water vapor (w_p) for different temperatures and an atmospheric pressure of 1000 mB. The bar marked (1ns) represents the variation of the refraction index which causes a propagation delay change of 1 ns, for a distance of 16 km.

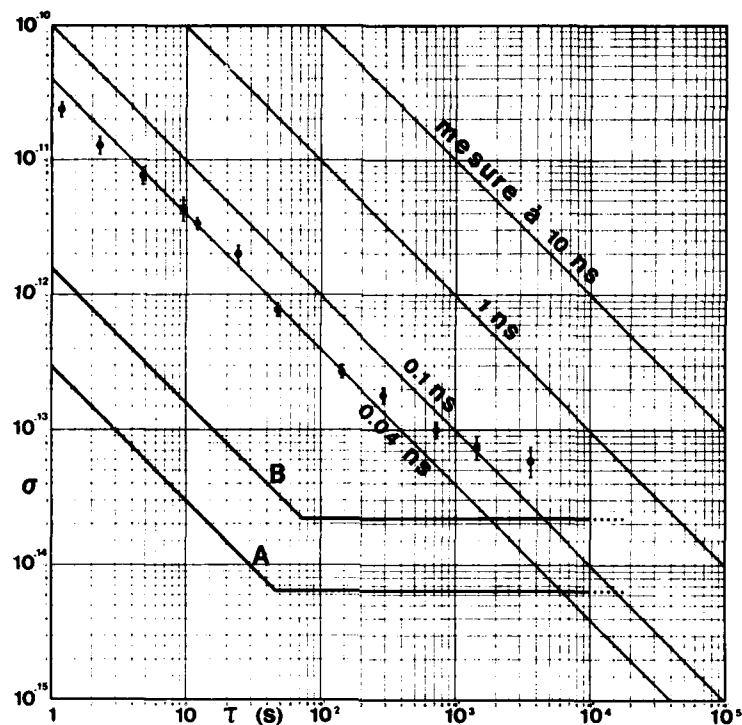


Figure 6- Results of the frequency stabilities measurements between distant hydrogen masers. The dots represent the experimental results which are consistent with a precision of 40 ps in the time comparison. Curves A and B represent the frequency stability of the signals delivered by the masers used at the 1.4 GHz output and the 5 MHz output respectively.

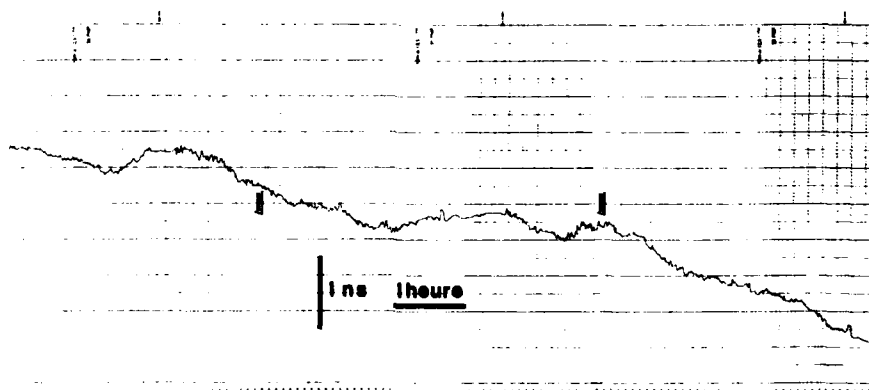


Figure 7. Phase record between two distant hydrogen masers.

STUDY OF L.F. AND V.L.F. TIME SIGNALS BY DIGITAL METHOD

F. Guillaume, J.C. Liévin and J. De Prins

Laboratoire des Etalons de Fréquence,
Université Libre de Bruxelles, Belgium.

Abstract

A quasi on-line data processing method is proposed for time comparison and time synchronization by means of L.F. and V.L.F. signals. The phase and amplitude of the received signal is measured and recorded in a sufficiently short time to allow the use of signal averaging techniques that greatly improve the S/N ratio.

Results on LORAN C, M.S.F. (A.M.) and G.B.R. (FSK) timing signals are presented. Special emphasis is placed on LORAN C signals. In this case, deconvolution techniques may be applied to practically restore the signal as it is received at the antenna. Absolute E.C.D. may thus be interpreted in terms of propagation effects. Some examples of decomposition into ground wave and sky wave components are given.

Key words : LORAN C, Time synchronization, Signal averaging, Deconvolution, Ground wave, Sky wave, F.S.K.

I. Introduction

Our laboratory is concerned with time comparison and time synchronization by means of L.F. and V.L.F. signals. We are also interested in sky wave propagation.

Three types of L.F. and V.L.F. timing signals were studied :

1. LORAN C, a pulsed signal with a spectral bandwidth usually narrower than the reception bandwidth. This very interesting feature enables us to introduce deconvolution techniques for the pulse analysis.
2. M.S.F., an A.M. transmitter. Here, the reception bandwidth is smaller than the signal bandwidth. Synchronization difficulties arise from the fact that the pulse shape is strongly dependent on local antenna conditions.
3. G.B.R., an F.S.K. transmitter.

Since results on M.S.F. and G.B.R. are already published elsewhere (1), (2), (3), only a brief outline of these techniques will be reported here, and emphasis will be placed on the study of LORAN C signals.

Studies of synchronization signals are often performed by means of analogic devices. These procedures are generally cumbersome and rather slow. A quasi on-line data processing method was developed to measure and to record in a few μ s the phase and the amplitude of the received signal. Furthermore, during the experiment, a check-up of all the characteristics of the receiver is performed. According to the results of this operation, a correction is introduced on the signal. In the most favorable case (LORAN C), the use of deconvolution techniques enables us to practically restore the signal as it is received at the antenna.

II. Need of deconvolution

It is well known that the effect of a filter on an

R.F. pulse is to produce on the carrier a delay depending mainly on the phase characteristics of the filter at the tuned frequency. On the other hand, the shape of the envelope will also be affected. As long as the envelope deformation is small, we can roughly consider this as a mere delaying of the envelope. It is clear that both delays - on the envelope and on the carrier - are in general different.

Hence, any attempt to interpret absolute E.C.D. in terms of propagation effects is meaningless when performed directly on the output signal. This signal is indeed the result of a convolution between the input signal, i.e. the LORAN C pulse as it is received at the antenna, and the impulse response of the whole receiving system, acting as a filter. Thus, actual measurements on the output signal result from the past history of the input signal at the measurement time, that, because of the delay introduced by the filter, already covers a much longer span.

As an example (Figure 1), let us assume that the delay introduced by the filter is 25 μ s. The determination of the 30 μ s tracking point on the output signal will in fact depend on the events that occurred during the first 55 μ s (30 + 25 μ s) on the input signal. This means that when the experimenter is measuring the tracking point, the received signal is almost reaching its maximum. Consequently, at the output, an assumed purely ground wave portion of the pulse may already be slightly contaminated by a sky wave. It follows that a complete understanding of the received LORAN C signal and its separation into ground wave and sky wave components requires the knowledge of the signal at the antenna. This can only be done by means of deconvolution techniques.

III. Experimental setup

A block diagram of the system is shown in Figure 2. The experimental setup is composed by a reception system for the LORAN C signals, a data acquisition and data processing system and a timing unit :

Reception system

The LORAN C signals are received by a loop antenna, consisting of a 7 turn coil of 72 cm diameter. This antenna has an impedance of 50 Ω at 100 kHz. The receiver is a 4 pole bandpass Butterworth filter, tuned to 100 kHz with a 20 kHz bandwidth. It is composed by four BURR-BROWN UAF 21 active filters, used as specified by the manufacturer (4). A low noise preamplifier matches the 50 Ω coaxial line of the antenna to the 100 k Ω input impedance of the receiver. The main characteristics of the receiving system are summarized in Figure 3.

Data acquisition and data processing system

The sample and hold DATEL SIM-UI has a 200 ns aperture time and an acquisition time of 35 ns for a 5 V input step. The analog to digital converter DATEL G8B3C is of the 8 bit type with a 800 ns conversion time. Control of the data and the signal

averaging are performed on-line by a 8 K VARIAN 520 i mini-computer. The D.M.A. has a 4.4×10^5 bytes/s transfer rate. The preprocessed data are sent by a telex line to a C.D.C. 6600 computer for complete processing.

Timing system

The triggering pulses for the sampling procedure are generated by a timing unit including a RHODE and SCHWARTZ X.S.D. quartz oscillator, synthesizers and an epoch monitor. The timing unit also produces pulses of 60 ns duration for the measurement of the impulse response of the antenna-receiver system. This 60 ns pulse is emitted by the auxiliary loop (Figure 2) at the G.R.P., between two groups of LORAN C pulses.

IV. Data processing

A preprocessing of the data is first performed on-line on the VARIAN 520 i minicomputer. Next, a more complete processing follows the sending of the preprocessed data to the C.D.C. 6600 computer.

The on-line processing of the data on the minicomputer involves the following steps :

1. Sampling of the pulse for an interval of 640 μ s at a frequency rate of 400 kHz.
2. Checking the reliability of the sampled pulse. Pulses not meeting predetermined criteria for maximum allowable deformation and noise levels are rejected.
3. Signal averaging.

Steps 1 to 3 are repeated until the specified number of pulses - usually 400 to 1000 - are averaged.

For monitoring purposes, the averaged pulse and the last sampled pulse are displayed on a scope. Figures 4 to 7 show typical displays, obtained from different stations of the Norwegian and the Mediterranean chains.

The on-line processing lasts about 70 ms. As the Norwegian and the Mediterranean G.R.P. are 79.7 and 79.9 ms, respectively, only the first pulse of the signal is analyzed.

The position of the synchronization pulse for the sampling procedure is adjusted so that the LORAN C pulse, as well as the impulse response, are centered on the middle of the sampling window.

The averaged data are sent by a telex line to a C.D.C. 6600 computer for complete data processing, that includes the deconvolution and the decomposition into ground wave and sky wave components. The deconvolution procedure is illustrated in Figure 8 : the averaged pulse, $p(t)$, and the averaged impulse response of the receiving system, $h(t)$, are Fourier transformed and deconvolved. The result of the deconvolution yields the Fourier transform $F(\omega)$ of the input signal. The LORAN C pulse, $f(t)$, is then easily obtained by taking the inverse Fourier transform of $F(\omega)$.

The spectrum of the theoretical LORAN C pulse, $F_t(\omega)$ and the spectrum of the receiving system, $H(\omega)$ are compared in Figure 9. As $F_t(\omega)$ is going to zero

much more rapidly than $H(\omega)$, $f(t)$ may be reconstructed with a minimum of distortion. In practice, the quality of the signal reconstruction will be limited by the noise on $H(\omega)$. Hence, to avoid degradation of the LORAN C signal by the noise on $H(\omega)$, the transfer function bandwidth has to be limited for the reconstruction. Thus far, $H(\omega)$ is usually defined by means of 25 spectral lines, ranging from 81.25 kHz to 118.75 kHz. Figure 10 a shows good agreement between a reconstructed theoretical LORAN C pulse after reduction of the transfer function bandwidth and the theoretical pulse itself. From Figure 10 b, it can be seen that the discrepancy in time between the two pulses does not exceed 0.75 μ s. Moreover, as the error introduced by the limitation on the transfer function bandwidth is known, it is possible to take it into account in further calculations.

Once the LORAN C pulse as it is received at the antenna is obtained, we can proceed to the separation into ground wave and sky wave components. This decomposition is performed by means of successive subtractions of the ideal shape in $t^2 e^{-\alpha t} \sin(\omega t)$ (5) from the corrected LORAN C signal. The parameter α is adjusted by minimizing the discrepancy between the ideal leading edge $t^2 e^{-\alpha t}$ and the leading edge of the pulse. For the LORAN C received at Brussels, α has a typical value of 2/62.5 μ s. In our analysis, the subtraction procedure is stopped when the amplitude of the remaining signal is less than 20% of the initial amplitude. In the future, more objective criteria need to be defined to stop the subtraction procedure.

V. LORAN C results

The ground wave and the different sky wave components are each characterized by their delay t_d , phase φ_M and amplitude A_r . The delay is the time interval elapsed between the beginning of the sampling window and the detection of each component. The phase is calculated at the zero crossing times of the carrier, around the maximum amplitude. The sky wave amplitudes are expressed as a percentage of the ground wave amplitude and are thus relative amplitudes. In Table I, r.m.s. values of these characteristics are given for Sylt, situated at 514 km from Brussels, and refer to results obtained since the beginning of this year.

Table 1

	Ground wave	Sky wave
Delay t_d :	0.80 μ s	1.46 μ s
Phase jitter φ_M :	15 ns	100 ns
Amplitude A_r :	2 %	4 %

Since Sylt is the LORAN C emitter nearest to Brussels, most results reported here refer to this station.

A typical example of decomposition into ground wave and sky wave is illustrated in Figure 11. This example refers to a pulse from Sylt, obtained by averaging over 400 valid pulses. The deconvolved pulse is shown above. In the middle, the ground wave is given with a normalized amplitude of 1. The detected sky wave, shown below, has a relative amplitude of 0.74 and a delay of 57.66 μ s with respect to the ground wave. This delay is consistent with a reflection between 60 and 65 km altitude.

The decomposition process depends on the knowledge of the exact pulse shape at the emitter. It is unfortunate that very little information is available in the literature. For instance, we do not know what the phase variation is during the decaying of the pulse.

However, the decomposition method may be applied in more complicated situations, where multiple sky waves are present, as shown in Figure 12. In this pulse, received from Sylt on March 6, 1977 at 23⁰⁰, three different sky waves were detected. The validity of the decomposition may be determined by the constancy of the phase in the last residual impulsions. Such a configuration can be explained by single and multiple reflections at 60 and 90 km altitude.

VI. M.S.F. and G.B.R. results

The same experimental setup developed for the LORAN C signals may be used for the study of A.M. and F.S.K. time signals. In these cases, the signal bandwidth is larger than the receiver bandwidth. Hence, deconvolution techniques may no longer be applied. Therefore, appropriate methods for the analysis of A.M. and F.S.K. signals were developed.

1. A.M. signals (M.S.F.)

In this case, the pulse shape is strongly dependent on local antenna conditions. It is also influenced by the variations of the transmitter characteristics. However, reproducible measurements may be obtained by defining a "characteristic point" at 85% of the mean amplitude (6). During 24-hour intervals, the r.m.s. error due to propagation and emission conditions is of the order of 20 μ s. Fluctuations of the order of 50 μ s are observed during one-year periods.

2. F.S.K. signals (G.B.R.)

F.S.K. is obtained by a commutation between 16 kHz and 15.95 kHz (7). The commutation is made when the phases of the two frequencies are both zero. This happens every 50 ms.

Due to the short distance (420 km) that separates our laboratory from the V.L.F. transmitter, interferences between ground wave and sky waves occur. Consequently, the results may be difficult to interpret (8). During day-time, the r.m.s. error obtained on the measurement of the time keying is of the order of 20 μ s although the F.S.K. emission from G.B.R. is not primarily intended for time dissemination (3).

VII. Conclusions

The digital method presently described is particularly well suited for time synchronization. Averaging and processing indeed permit a better precision. Large volumes of information can be very easily processed and programming versatility allows, with the same equipment, measurements on signals as different as A.M. and LORAN C. In the future, once the procedure is made operational, a continuous and automatic processing would be possible with the help of a microprocessor.

ACKNOWLEDGMENTS

The authors are grateful to Dr L. Bossy, from the Université de Louvain, Belgium for his advice in the interpretation of delays between the different sky

waves in terms of propagation effects. They are also pleased to thank Dr R. Coutrez and Dr W. Scholiers, from the Université Libre de Bruxelles, Belgium for their interesting discussions on the LORAN C pulse shape. Finally, they appreciate the help of their colleagues, Dr J.L. Guisset, Dr G. Cornélissen, R. Detrie and J.P. Lechien during the realization of the experiments.

References

- (1) J. De Prins, "Timing Systems", E.L.F. - V.L.F. Radio Wave Propagation, edited by J.A. Holtet - NATO Advanced Study Institutes Series C, pp. 335-398, 1974.
- (2) J. De Prins, R. Detrie, J.L. Guisset, F. Guillaume, "Time Pulse Measurement by Digital Method", Proc. of the second Cagliari International Meeting on Time Determination, Dissemination and Synchronization, pp. 151-164, 1975.
- (3) "Etude par Méthode Digitale des Signaux Horaires à B.F. et T.B.F.", Document Commission d'étude du CCIR, période 1974-1978, document 7/8-F, 20 octobre 1975.
- (4) Burr-Brown, Application Note 61, July 1973.
- (5) C.E. Potts, B. Wieder, "Precise Time and Frequency Dissemination via the LORAN C System", Proc. IEEE 60, 5, pp. 530-539, May 1972.
- (6) D.H. Andrews, C. Chaslain, J. De Prins, "Reception of Low Frequency Time Signals", Frequency 6, 9, pp. 13-21, 1968.
- (7) R.R. Stone, T.H. Gatus, T.N. Lieberman, "Utilization of F.S.K.", Time Proceeding of the Fourth Annual NASA and Department of Defense - Precise Time and Time Interval Planning Meeting, pp. 324-344, 1973.
- (8) B.E. Blair, "Time and Frequency Dissemination, an Overview of Principles and Techniques", ch. 19 of N.B.S. Monograph 140, pp. 233-313, May 1974.

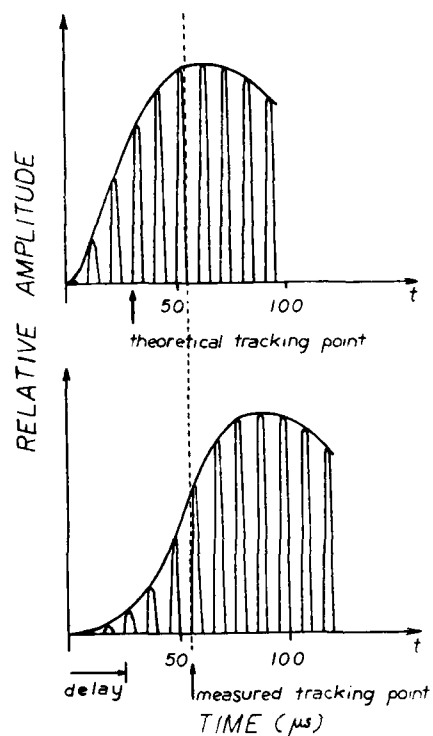


Figure 1. Theoretical LORAN C pulse at the antenna (above) and at the output of a receiving system that produces only a delaying of the envelope (below).

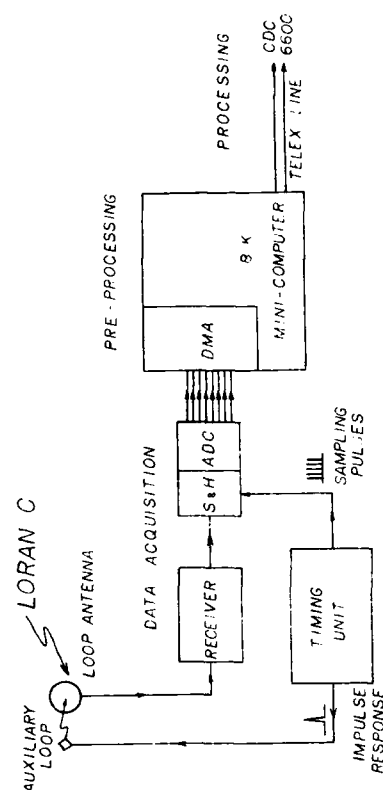


Figure 2. Block diagram of the experimental setup.

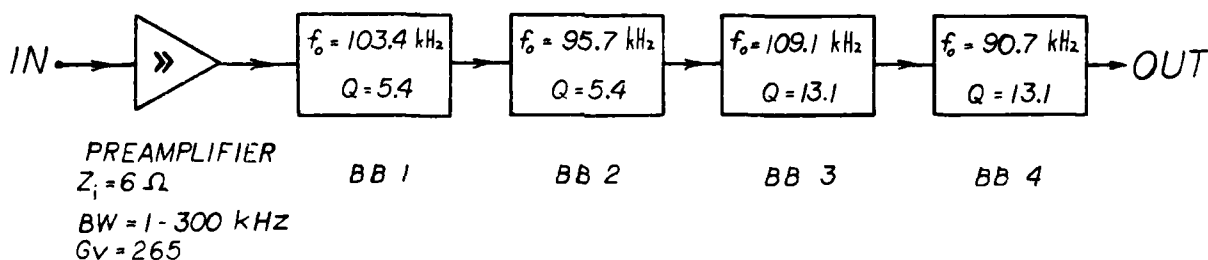


Figure 3. Main characteristics of the preamplifier and the Butterworth active filter. All stages of the filter have the same gain of 5. The central frequency f_0 and the quality factor Q are indicated for each stage. The voltage gain of the preamplifier & receiver system is 15900.

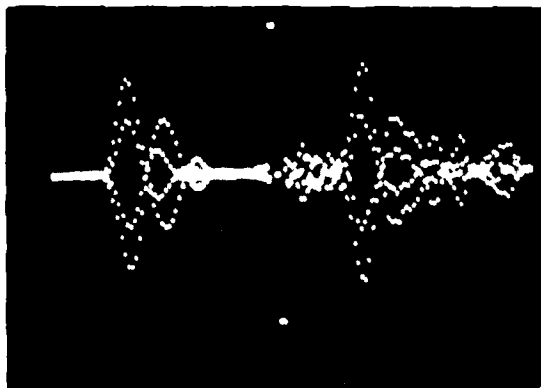


FIGURE 4

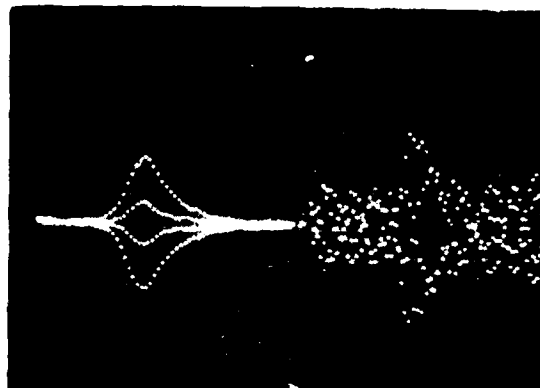


FIGURE 5

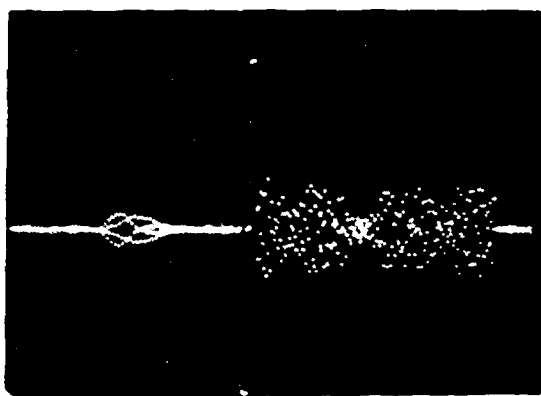


FIGURE 6

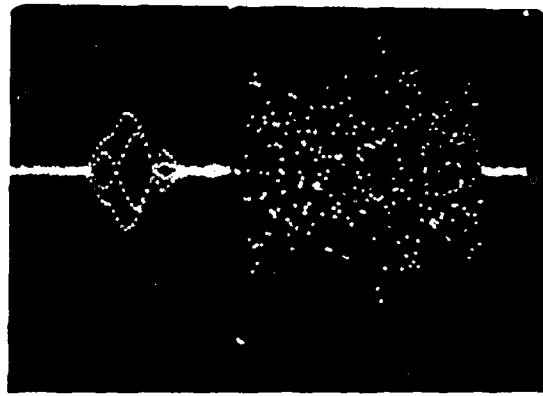


FIGURE 7

Figures 4 to 7.

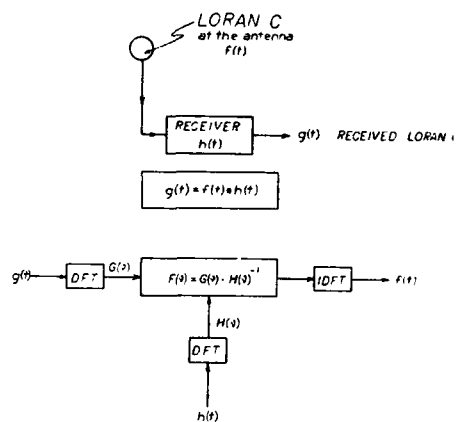
These figures show different LORAN C pulses of the Norwegian and the Mediterranean chains, after the sampling and the averaging. The right part of each figure is the pulse as it is sampled. The result of the averaging of the last pulse with the previous ones is shown on the left part of the figure.

Figure 4 : SYLT (Norwegian chain - 514 km from Brussels)

Figure 5 : EJDE (Master of the Norwegian chain - 1450 km from Brussels)

Figure 6 : SANDUR (Norwegian chain - 2200 km from Brussels)

Figure 7 : LAMPEDUSA (Mediterranean chain - 1820 km from Brussels)



DECONVOLUTION

FIGURE 8. Deconvolution procedure. $h(t)$ is the impulse response of the antenna-receiver system.

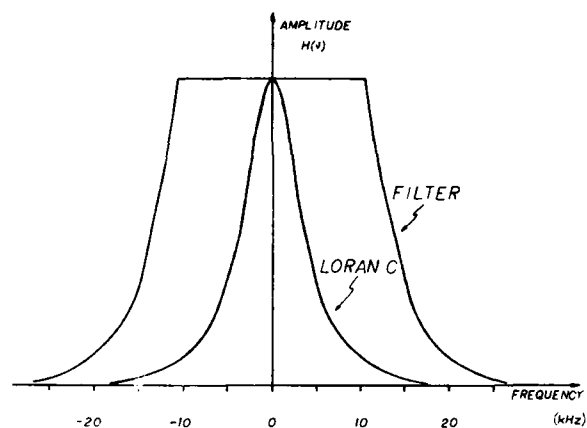


FIGURE 9. Comparison of the transfer function $H(s)$ of the receiving system and the spectrum of a theoretical LORAN C pulse.

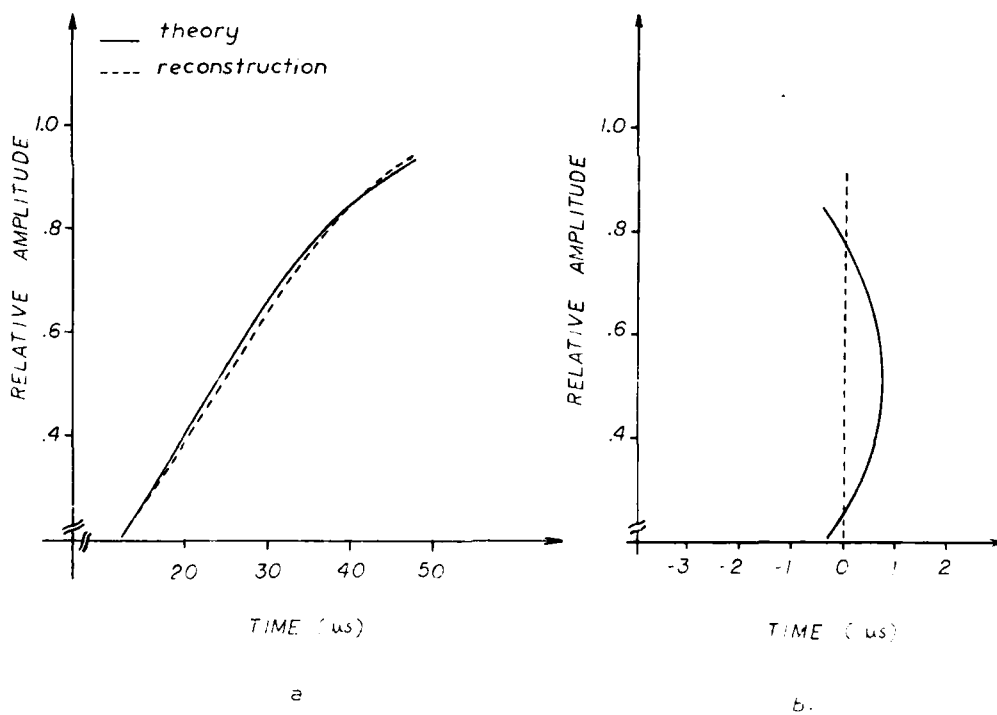


FIGURE 10.

- 10a. Good agreement between the leading edge of a theoretical LORAN C pulse and its reconstruction with 25 spectral lines.
10b. Discrepancy in time between the theoretical and the reconstructed leading edge.

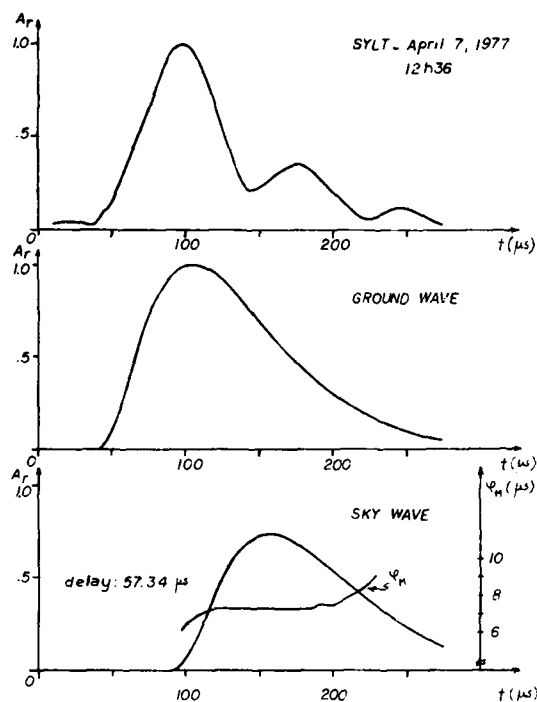


FIGURE 11. Decomposition into ground wave and sky wave of a pulse received from Sylt on April 7, 1977 at 12h36.

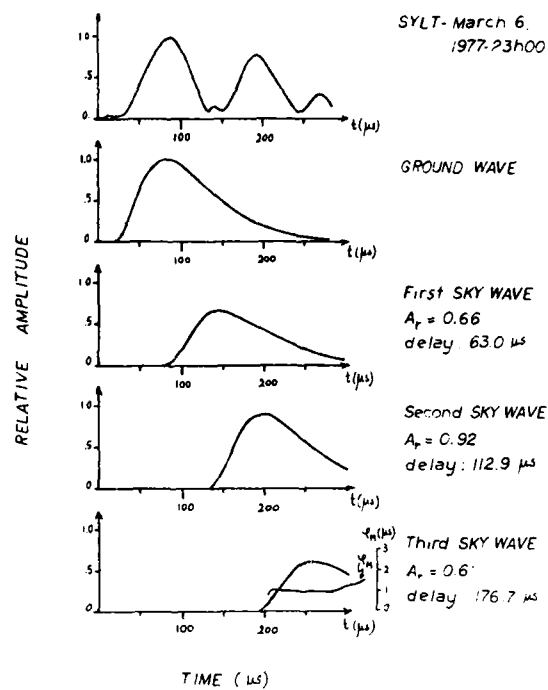


FIGURE 12. Decomposition into ground wave and sky waves of a pulse received from Sylt on March 6, 1977 at 23h00.

DESIGN OF A SPACECRAFT HYDROGEN MASER*

Edward S. Sabisky
RCA David Sarnoff Research Center
Princeton, N.J. 08540

and

H. E. Peters
Consultant, Previously with
NASA Goddard Space Craft Center
Greenbelt, Maryland

Introduction

Under a program to develop a very stable clock for the NAVSTAR Global Positioning System (GPS) satellite, several designs using the hydrogen atom, have been considered. These include the passive resonance and the active oscillator as well as various cavity mode and storage cell configurations. A design has been chosen which uses the hydrogen maser principle. The philosophy has been to achieve a high performance in frequency stability and, where possible, to use design approaches which have been proven by experience with long term operation in ground based hydrogen masers. This approach is feasible while still satisfying the size, weight and power restriction required by space craft. The main objectives are to achieve a stability of better than 1 part in 10^{14} for averaging times between 100 seconds and 10 days and to have a weight less than 90 pounds within a space of 15 inches (38 cm) diameter by 30 inches (76 cm) long. In this paper, the design of the spacecraft hydrogen maser will be presented and test data on completed components will be given.

Overall Description, Design and Performance

The atomic hydrogen maser⁽¹⁾ is a conceptually simple device as shown by the block diagram of Fig. 1. Molecular hydrogen is purified and enters a glass chamber where it is dissociated into atoms of hydrogen by an RF discharge. Neutral atoms of hydrogen emerge from the glass discharge chamber through a small aperture (collimator) and then pass through a magnetic state selector. In the state selector, undesired atoms are defocused while the desired atoms (atoms in the two higher energy states) are focused to pass into a large storage cell which is located in a microwave cavity. Due to maser action, the atoms in decaying to a lower energy state emit radiation at a well defined frequency near 1.4 GHz which is then coupled out of the cavity and can then be processed to match the user requirements. This is a simplified version of the clock and an actual hydrogen clock must include magnetic and thermal shielding, magnet field coils, vacuum enclosure, pumps and various electronic control systems.

The design presented in this paper is primarily intended for an advanced development model of a clock for use in the Global Positioning System. Our principal philosophy is to build a reliable, long-lived clock having a stability of at least 1 part in 10^{14} while meeting the stringent environmental requirements for space travel. In order to meet the system requirements, the weight of the clock must be less than 90 pounds and must fit within a volume of 38 cm (15 inches) in diameter by 76 cm (30 inches) long. The design is largely based on research results of NASA Goddard⁽²⁾. A cross sectional view of our design is shown in Fig-

ure 2. The details of the various component parts of the clock are given in a later section. The basic features of our design are:

- Standard low pressure vessel for storage of molecular hydrogen;
- A palladium-silver purifier and valve;
- A capacitively RF coupled dissociator using a pyrex bulb having an ion-depleted inside surface;
- The collimator consists of a single 0.0127 cm diameter hole in 0.1016 cm thick glass wafer;
- A small quadrupole state selector;
- A cylindrical storage cell of thin film FEP teflon;
- A microwave cavity consisting of a molybdenum cylinder with aluminum end plates;
- Fine tuning of cavity done by thermal expansion;
- Magnetic field coil with trim coils on both ends;
- Two vacuum chambers;
- Four magnetic shields;
- Two secondary temperature isolation ovens;
- Degaussing capability.

With the design features given above, the theoretical performance parameters of the development model of the space clock are:

(a) Line Q_L	5.5×10^9
(b) Loaded cavity Q_c	5.5×10^4
(c) Cavity pulling factor $\frac{Q_c}{Q_L}$	1×10^{-5}
(d) Beam threshold flux for oscillation	$1.9 \times 10^{11} \frac{\text{atoms}}{\text{sec.}}$
(e) Net beam flux $\frac{I}{I_{th}} = 6$	$9.6 \times 10^{11} \frac{\text{atoms}}{\text{sec.}}$
(f) Power delivered by H-atoms at $\frac{I}{I_{th}} = 6$	$5.3 \times 10^{-13} \text{ watts}$
(g) Short term stability $\frac{\Delta f}{f}$ ($0.1 \leq \tau \leq 10^2$)	$2.0 \times 10^{-13} / \tau$, τ in sec.
(h) Long term stability ($10^2 \leq \tau \leq 10^6$)	$\frac{\Delta f}{f} < 10^{-14}$
(i) Perturbing noise level $\frac{\Delta f}{f}$ ($10^2 \leq \tau \leq 10^4$)	$9.0 \times 10^{-15} / \sqrt{\tau}$
(j) Spin exchange parameter "q"	0.05
(k) Transverse shielding factor	130,000
(l) Axial shielding factor: passive active	30,000 > 10^5
(m) Molecular hydrogen consumption	0.02 moles/year

Apart from self-explanatory terms, the parameters above are: I is the net beam flux (the difference in the flux of atoms entering in the higher energy state of the maser transition and the lower energy state); I_{th} is the threshold flux for oscillation to occur provided-

* Work supported by the office of Naval Research under Contract N00014-75-C-1148.

ing spin exchange is neglected; q is a quality factor of the clock, see D. Kleppner; et al⁽³⁾ for further definitions. The predicted performance of the clock is given in Figure 3.

Component Design Description

Molecular Hydrogen Supply

The hydrogen supply in all ground-based hydrogen clocks is a cylinder with a volume of about 1-liter and the gas at a pressure of 1000 psi. In the space probe maser⁽⁴⁾ developed by the Smithsonian Astrophysical Observatory (SAO), the hydrogen supply consisted of about 70 grams of LiAlH_4 . When heated, the LiAlH_4 releases hydrogen at a rate controlled by the temperature of the material. In this design, we intend to use a pressure vessel with a volume of 0.4 liter and the gas at 300 psi. This corresponds to a storage of 0.4 mole of molecular hydrogen and with a hydrogen consumption rate of 0.02 moles per year, the hydrogen would last 20 years. This type of storage is simple and reliable and does not require any preparation of new materials. The container weighs about 453 grams (1 pound) and occupies a space of 500 cm^3 . Thus the only disadvantage in the above approach is size. If the size becomes critical, a metal hydride can be used for the storage of hydrogen. There are a number of metal hydrides⁽⁵⁾ which can be used and some testing is needed before final selection. The candidates are depleted uranium metal, the $\text{LaNi}_5\text{-H}$ system and LiAlH_4 . For example, 100 grams of uranium metal would adequately supply the needed molecular hydrogen while occupying a space of 5-10 cm^3 . For the depleted uranium metal, the pressure $P(\text{mm})$ of the hydrogen at a temperature T is given by

$$P = 1.6 \times 10^9 \exp \left[-1.027 \times \frac{10^4}{T} \right]$$

Purifier/Valve

The purifier to be used consists of a solid metallic pellet alloy of 75% palladium and 25% silver which restricts gas flow through a small tube. The pellet permits hydrogen diffusion while blocking all other matter and the rate of hydrogen diffusion is controlled by the temperature of the alloy. The mechanical assembly shown in Figure 4 is relatively easy to fabricate and process. The incorporation of the copper sleeve improves thermal conductivity and welding. The copper-stainless steel joints need no special preparation for brazing and the copper-copper and copper-pellet surfaces can be very readily brazed together. The temperature of the pellet is varied by a resistance heater wound on the copper sleeve. A temperature of about 75°C is needed to achieve a pressure of 0.05mm in the dissociator as shown in Figure 5.

Pressure Control

The pressure control/gauge is simply a thermistor having a small thermal inertia. The thermistor sensor is placed in a copper block whose temperature is held constant by a separate sensor which controls a heater wound around the copper block, thus ensuring that the pressure sensor is in a constant temperature environment. The pressure sensor is part of a bridge circuit in which the emf of the sensor is compared to an adjustable reference voltage in order to establish the heating current for the palladium-silver leak and thus maintaining the source pressure at the desired level. The response of the pressure regulator to steps in the reference voltage is shown in Figure 6.

Dissociator and Collimator

Molecular hydrogen is converted to atomic hydrogen in the dissociator. The most popular form of dissociator is a glass chamber in which a gaseous discharge is maintained. The discharge is driven by an R.F. field having a frequency near 110 MHz. The radio frequency is coupled to the plasma either inductively or capacitively using external electrodes. Molecular hydrogen is dissociated by collisions with hot electrons in the plasma. The hydrogen atoms formed diffuse towards the wall where they may recombine. Due to the low probability of recombination, a high density of atomic hydrogen is built up in the discharge chamber. Hydrogen at a steady state pressure, typically 0.05mm, is maintained in the discharge chamber by suitably controlling the palladium leak valve. The atomic hydrogen beam exits through a small hole or holes (collimator) in the wall of the glass chamber.

Both the discharge condition and the surface condition of the wall are important in order to optimize the dissociator performance. The discharge determines the power requirements to obtain a certain dissociation rate. The wall determines the recombination rate, which critically depends on wall temperature and on wall contamination. The discharge plasma and the wall interact and thus optimization must involve consideration of both.

A sketch of the dissociator design and its metal housing are shown in Figure 7. The glass bulb is typically 6.3 cm (2.5 inch) in diameter and 7.1 cm (2.8 inch) long. The frequency is around 110 MHz, the total dc power required for the electronics is 11 watts, and capacitive coupling is used. We have three complete dissociator units made from different glasses for the bulb under life test. The three bulbs are: pyrex 7740, fused quartz, and ion-depleted pyrex 7740. A photograph of an operating unit is shown in Figure 8.

The dissociator life test system with the pyrex bulb has been operating continuously since August 1976 with no measurable degradation in performance and negligible coloring of the glass. The test units with quartz and the ion-depleted glass bulb have been operating continuously since February 1977. [Ion-depletion is a process that removes mobile ions from the surface of a glass by the application of an electric field to a "blocking" anode at temperatures of a few hundred degrees centigrade. A "blocking" anode is one that does not inject ions into the glass and this anode must allow oxygen to evolve from the glass surface.] Using optical detection of the atomic hydrogen lines as a measurement tool, we find only small differences in the intensity of the atomic hydrogen lines between the three units. The temperature of the pyrex bulbs without any cooling is 90°C while that of the quartz bulb is 98°C. However, the electronic circuit was optimized only for the pyrex bulbs. Both pyrex bulbs ignite without the use of a separate excitation circuit. The quartz bulb ignites but is more sensitive than the pyrex bulbs. From spectroscopic tests, we find that the optical intensity of the atomic hydrogen lines are substantially more intense than the molecular hydrogen lines (see Fig. 9). We have found that increasing the pressure causes the atomic hydrogen spectral intensity to decrease and increasing the temperature of the bulb causes a decrease in the atomic hydrogen spectral intensity.

We intend to use a pyrex bulb in our demonstration model with an ion-depleted surface. The dissociator assembly will include an optical monitor for optical monitoring of the atomic hydrogen Balmer α -line 6563Å. The collimator will consist of a single 0.0127 cm diameter hole in one end of the glass bulb.

State Selector

The maser criterion requires that the number of atoms in the upper energy state be larger than the number of atoms in the lower energy state. This is an abnormal condition and certain techniques must be employed to reach this non-equilibrium population distribution. The hydrogen atoms as they emerge from the dissociator bulb would have a normal Boltzmann population distribution of atoms, i.e., the lowest energy states would have the largest number of atoms. Because the operating temperature is high, the four lowest energy states are essentially equal in population. The function of the state selector is therefore to filter out the lowest energy state atoms and allow the atoms in the higher energy states to reach the cavity. Thus, the state selector maximizes the flux of the atoms in the upper state delivered to the cavity while blocking or minimizing the number of lower state atoms reaching the cavity.

In the past, two types of magnet designs were employed; the hexapole design which has a quadratic variation of field strength with radius from axis and the quadrupole design which exhibits a linear variation of field strength with radius. Both types of units have been successfully used in operating ground based clocks. The design of the units have also improved with time and are now more efficient than the older models.

From analysis and from the experiments of H. Peters at NASA GSFC with state selectors of differing lengths, bore diameters, field strength and with and without stopping discs; the quadrupole design was shown to have an advantage over the hexapole design. For example, the quadrupole does not require a stop on the beam axis as required by the hexapole to interrupt on-axis atoms with zero transverse velocity. The transverse magnetic deflection force exerted by the hexapole is zero on axis while that of the quadrupole is constant with radius. The quadrupole design has greater gaps between poles and provides higher pumping speeds for the unfocused hydrogen. The quadrupole design is simpler and easier to make--four versus six poles.

We have designed a small quadrupole state selector which will be used in the maser now under construction. The design is simple and employs soft iron for the cylinder and for the pole tips (see Figure 10). Four rectangular permanent magnet pieces are placed between the soft iron cylinder and tips. The permanent magnetic material is rare earth-cobalt, a standard item which can readily be purchased. The material has an energy BH product of 12 million gauss-oersted and the design will give a minimum of 10 kilogauss at the maximum field position. The unit has a diameter of 3.8 cm and a length of 3.2 cm with a bore diameter of 0.05 cm. The weight of the unit is 0.8 pound (362 grams) and occupies a volume of 36 Cm³. This design of the state selector permits a distance of 8.25 cm between the state selector exit and the entrance to the cavity.

Magnetic Field and Magnetic Shielding

The ground state energy levels of atomic hydrogen are such that the two lowest energy states characterized by the quantum labels F=1 and F=0 are not degenerate even at zero magnetic field. The ground state energy levels have a zero field splitting. It is possible therefore to theoretically operate a hydrogen maser in zero magnetic field. This raises the practical problem of achieving essentially zero field which is exceedingly more difficult than operating at some finite magnetic field. Therefore, masers operate with applied magnetic fields which are somewhat larger than the background noise level. Practical operating masers

have an applied field of 0.5 to 1.0 milligauss. The applied field is normally generated by passing current through a solenoid whose axis coincides with that of a TE₀₁₁ cylindrical cavity. The stability requirements of the current source is a function of the applied field strength. For example for a stability of 1 part in 10¹⁵ at a field of 10⁻³G, the required current stability is 0.025% while at an operating field of 10⁻⁴ G the required stability is only 2.5 %. Thus the stability of the current supply for the solenoid is not a problem. Furthermore, the lower the applied field, the smaller is the effect of the field (current) fluctuations on the clock frequency. This is seen from the relationship between the maser frequency f, and the magnetic field H, given by

$$f = f_0 + 2.75 \times 10^3 H^2$$

where f₀ = zero field hyperfine splitting frequency in Hz and H is the applied magnetic field in Gauss.

A more difficult problem arises due to the fluctuations of the ambient magnetic field if not properly shielded. The earth's field is typically 0.5 gauss and urban magnetic noise in the frequency range of 0.1 to 100 Hz is typically 10⁻² gauss. There are stray magnetic fields from components such as Vac-Ion pumps and some external electrical equipment are also a source of magnetic field fluctuation. On ground based clocks, magnetic shielding has been accomplished by using static ferromagnetic shielding. Three nested shields have been used, four, five and six shields. Shields of various thicknesses and shape factors have also been employed. The processing of the material is very important and there always remains some remnant magnetization in certain parts of the material. Although the magnetic shielding-problem is difficult for ground-based clocks, it is grossly exaggerated in clocks for space application because of the weight factor. It would not be surprising if the weight of the magnetic shields were to constitute about 35% of the weight of the space maser.

In our analysis of the magnetic shielding of the advanced development model of the space clock, it was clear that a trade-off between weight and shielding is necessary. We have made calculations on a number of different configurations of shields. Our shielding design consists of four shields with each shield having a thickness of 0.063 cm (a standard thickness) and the material is 4-79 moly-permalloy.⁽⁶⁾ The transverse shielding factor is 130,000 and the axial shielding factor is 30,000 as shown in Figure 11. The operating maser would then have a calculated sensitivity to external magnetic field fluctuations of

$$\frac{\Delta f}{f_0} = \frac{5500}{f_0} \frac{H_z}{S_A} \Delta H_{\text{ext}}$$

where f₀ is the maser frequency; H_z is the applied magnetic field, S_A is the axial shielding factor, and ΔH_{ext} is the change in the external field. For H_z = 0.5 milligauss and S_A = 30,000, the sensitivity is 6.5x10⁻¹⁴ Gauss⁻¹. Thus static shielding alone without adding appreciably more shields (weight) would be adequate for variations of 0.1 gauss or less. Further improvement of the shielding and/or to reduce weight requires investigation of methods using dynamic shielding.

Storage Cell

The basic reason for the inherent higher frequency stability of the hydrogen clock over the cesium beam clock is that the line Q (frequency divided by resonance width) for the hydrogen clock is typically 100 times that of the cesium clock. A typical resen-

ance width of 250 Hz is obtained for the cesium beam while a one Hz width is usual for the hydrogen atoms. The high line Q or the narrow resonance line for hydrogen is due to the long interaction time of the hydrogen atoms with the microwave field. In the advanced development model of the clock, the storage cell has a volume of 6.6 liters. The atom enters the bulb with a velocity of 2.6×10^5 cm per second, bounces around approximately 20,000 times and after about a second departs. The average distance traveled between wall collisions is 12 cm and the hydrogen atom travels about a mile even though it is contained within the volume of a single mode cavity. In order to achieve a narrow resonance width or long lifetime for the hydrogen atoms, any perturbations of the atoms must be kept to a minimum. There are a number of effects which can be considered and the most important is that of the effect of wall collisions. The wall collisions can be divided into two categories: adiabatic and nonadiabatic. In an adiabatic collision the hydrogen atoms interact elastically with the walls and a small change in the energy levels (or frequency) always occurs. In the nonadiabatic, the atom is effectively lost either by a change of state or by chemical reaction with the surface of the wall. Thus the surface of the storage cell is of primary importance to the stability, accuracy and reproducibility of the atomic hydrogen clock.

In just about all of the present operational clocks, the storage cells are fabricated by making a cylindrical or a spherical container from fused quartz and then coating the inside surface with an appropriate coating. Measurements of the quality of the wall coatings are very difficult to make but it has been established by experiment that teflon-like materials are the best coatings and one form or another of teflon are used in all operating clocks. In our design, we intend to fabricate the storage cell from pure teflon film⁽⁷⁾ having a thickness of 0.0025 or 0.0051 cm. The film bulb storage cell has a number of significant advantages over the teflon coated quartz cell. 1) The teflon film is homogeneous and does not contain holes in the coating as one usually obtains when coating a quartz cell. The hydrogen atoms interacting with the quartz surface significantly reduces the lifetime. 2) From measurements⁽⁸⁾ made at NASA-GSFC, the hydrogen atoms are affected less by the film surface than by the coating deposited on the quartz. This means a smaller frequency shift or correction due to the reaction with the walls. 3) The cells made from the film should be very reproducible. 4) The film cell has a much smaller perturbation on the cavity frequency than the coated-quartz cell. The shift of the cavity frequency by a 0.0025 cm thick film has been measured to be 270 kHz compared to 60 MHz for a typical quartz cell. This means that the cavity, and therefore the clock frequency, is much less susceptible to changes in the cell caused by mechanical or temperature effects. 5) The teflon-film cell is naturally less susceptible to breakage than a quartz cell. 6) Since the frequency perturbation on the cavity is much smaller for the film-cell, the cavity need not be fabricated for each cell as is done for quartz cells. 7) The cost of the film cell is much cheaper and can be readily made without special glass blowing techniques. 8) The difficult step of uniformly coating the inside surface of a quartz cell is eliminated by using the film-cell.

Fabrication of Storage Cell. "Teflon" is the registered trademark of the DuPont Company for its fluorocarbon resins. FEP "Teflon" resins are tetra-fluoroethylene-hexafluoropropylene copolymers, usually fabricated by melt extrusion or injection molding. FEP type "A" film of 0.0025 and 0.0051 cm thickness was purchased from a DuPont Company distributor. The premium grade type "1" film is not available in thicknesses less than 0.0127 cm. An open-ended cylinder was

made using an overlap heat seal in the arrangement shown in Figure 12. In this fixture, the Teflon film is held between two pieces of fused quartz plates. Two nichrome heating strips are placed on the quartz plates and the strips are held in place by two aluminum plates which have one side bonded with synthetic mica 0.05 cm thick. Two weights of 24 lbs. each are used to apply pressure to the unit with a pressure of about one psi on the Teflon. A thermocouple (Chromel-Alumel) is placed midway in the fixture and about 0.6 cm from one of the heating strips. The thermo couple is used to monitor and establish the conditions necessary for a good seal. A simple procedure for obtaining good seals was developed. With the Teflon in place, the variac is turned up until the current meter reads 50 amperes and after about 50 seconds or more accurately when the voltage from the thermocouple read 3.1 mV ($\sim 70^\circ\text{C}$) the variac is turned down. In a separate experiment, it was established that the temperature at the Teflon joint was 257°C at the time the variac was turned down. If the temperature at the bond was lower, the seal was poor and easily came apart. If the temperature is too high, bubbles formed and the Teflon stuck to the quartz. The conditions for sealing were similar for both the 0.0025 and 0.0051 cm thick films.

Cell Assembly. A simple aluminum frame (see Figure 13) is made consisting of two aluminum rods and two aluminum plates. The inside dimensions of the frame should be large enough to enclose the cell and also to allow working space at both ends. Two aluminum end plates are fabricated which hold the Teflon film and eventually become part of the microwave cavity. A round aluminum rod is attached to each end plate to hold the cell to the frame. These rods are removed when the cell is attached to the cavity. Teflon film of 0.0025 cm thickness is placed over one side of the end plates and the film is brought to the opposite side, stretched and held in place by a metal ring. The open-ended cylinder of Teflon film is now placed over the end plates and both ends of the cylinder are worked carefully until the shape of the cylinder is satisfactory. The Teflon film at the cylinder ends is now clamped to the metal end plates by another metal ring. The first metal ring that was attached to hold the film covering the face of the end plates can now be removed and excess Teflon removed. The length of the cell should be some predetermined distance less than the final length to be used in the cavity. To maintain a rigid cylinder, the film-cylinder should always be in tension and about a 10% elongation of the length for a film thickness of 0.0025 cm is satisfactory. Measurements were made of the tensile stress in the film as a function of time after an initial stress of 1870 lbs per square inch (psi). The data indicates that the tension stress would be 800 psi after 10 years.

Escape from the Storage Cell. The stem (entrance collimator to storage cell) determines the lifetime of the atoms due to the probability of the atoms escaping from the cell. The escape rate of atoms from the cell γ_0 is found by equating the incident beam flux with the emergent flux $NKvA_e/4$ where N is the density, v is the mean velocity, A_e is the area of the stem and K is a factor depending on the geometry of the stem. If the volume of the storage cell is V_b , then the storage time T_b is determined by

$$T_b = \frac{4V_b}{vA_eK}$$

For a cell size V_b of 6.69 liters, $v = 2.6 \times 10^5$ cm per sec, $K = 0.1367$ and the diameter of the exit hole of 0.635 cm, $T_b = 2.4$ seconds. Therefore if all other relaxation mechanisms were not important, the lifetime of the hydrogen atoms in the cell would be limited by

the exit aperture and length of the stem and the lifetime for this design is 2.4 seconds. In which case, the maximum $Q_L = f\tau_L = 10^{10}$.

Microwave Cavity

The microwave cavity design is a critical component in the clock because: (1) the cavity must have a sufficiently high Q (low losses) in order to achieve maser operation; (2) the stability of the clock, particularly the long-term stability, is strongly coupled to the frequency stability of the microwave cavity-storage cell combination; (3) for space craft systems where size and weight are important factors, the size and weight of the clock is to a large extent determined by the size of the cavity.

There are two different modes of operation of the clock; namely, active and passive modes. Since the cavity requirements are very different for the two modes, and our design uses the active mode, this discussion on cavities will only be directed to the active mode. In this mode of operation, two different electrical designs of the cavity have been proposed and used. (1) A cylindrical cavity using the TE_{011} mode and (2) a cavity using the TE_{111} mode. The cavities designed for the TE_{011} mode have the higher Q (the unloaded Q for a copper surface is 80,000 for the TE_{011} compared to 33,000 for the TE_{111}). The TE_{111} mode cavity(9) in this application requires that a septum made from a material like teflon be positioned to divide the cavity into two parts. The septum is necessary because of the RF magnetic field configuration which makes it necessary that the hydrogen atoms in one-half of the cavity do not travel in the other half of the cavity. For an ideal system, the hydrogen atoms should be located in the cavity where the RF magnetic field direction is parallel to the applied DC magnetic field direction. The RF magnetic field configuration for the TE_{011} is more suited to the above requirement than the magnetic field configuration for the TE_{111} . The spin exchange parameter "q" of the clock is typically $q=0.05$ for masers using the TE_{011} mode cavity while a $q=0.16$ is expected for masers using the TE_{111} mode cavity. Since maser threshold requires a $q=0.172$, the masers using the TE_{111} mode cavities are not far from threshold and thus become very susceptible to ageing of material and environmental changes. The singular advantage of the TE_{111} mode cavity is its size which is inherently smaller than for the TE_{011} mode cavity. A typical size cavity for the TE_{111} mode is 15 cm in diameter by 15 cm long while for the TE_{011} mode the length is 28 cm and the diameter 28 cm. However, a smaller cavity means a smaller storage cell which means a larger wall shift and a lower line Q and therefore a reduced stability of the clock. For these reasons and particularly because of the many years of experience using the TE_{011} cavity compared with only very recent laboratory operation of a single clock with a TE_{111} mode cavity, the cavity design using the TE_{111} mode is not being seriously considered at this time. In addition, we feel that reducing the cavity size of the TE_{011} mode cavity by various techniques is more promising than the use of the TE_{111} mode cavity.

The long term stability of the clock is strongly correlated with the stability of the cavity-storage cell system. The problem is summarized by noting that the fraction of the atomic resonance linewidth by which the maser is "pulled" by a mistuned cavity is identical to the fraction of the cavity linewidth by which the cavity is mistuned. The shift of the oscillation frequency f from the true resonance frequency f_0 by a cavity tuned to a frequency f_c is

$$\frac{f-f_0}{f_0} = \frac{f_c-f_0}{f_0} \frac{Q_c}{Q_L}$$

where Q_c is the cavity Q and Q_L if the line Q of the atomic hydrogen. The smaller the ratio $\frac{Q_c}{Q_L}$, the less sensitive is the oscillator frequency to changes in the cavity frequency. Typically for a TE_{011} mode cavity $Q_c=50,000$ and $Q_L=5 \times 10^9$ or $\frac{Q_c}{Q_L} = 1 \times 10^{-5}$. In one case(9), a TE_{111} mode hydrogen maser resonant cavity was built and the measured values were $Q_c=27,000$ and $Q=8 \times 10^8$. The maximum line Q realizable should vary as the surface to volume ratio of the cell, which for typical TE_{111} mode designs is about one-half that of the TE_{011} design. Also, a shift in the cavity frequency of 1.4 Hz results in $\frac{\Delta f}{f_0} = 1 \times 10^{-14}$ for $\frac{Q_c}{Q_L} = 1 \times 10^{-5}$.

Two configurations of cavity design have been used and tested in the past. In one design(10), the cavity is made using a dielectric like Cer-vit which has a very small thermal coefficient of expansion and the inside surface is copper or silver plated. In the other design(11), the cavity is made from a metal such as aluminum which has a rather large thermal coefficient of expansion. The large coefficient of expansion is utilized for fine tuning of the cavity. In Table I we give the properties of the cavity and the storage cell as they relate to the temperature stability requirements of the cavity structure to achieve a clock stability of at least 1×10^{-14} . Based only on this criterion, one should use a cavity made from a material with a small thermal coefficient of expansion such as Cer-vit combined with a very thin teflon storage cell. However, the stability of the cavity due to temperature changes is only one of a number of factors to be considered and evaluated. The mechanical design is very important and a difficult to measure parameter like the dimensional stability(12) of the material being used must be evaluated.

There are a number of factors relating to the use of dielectric cavities which can be identified. (1) Copper or silver plating of about 0.0025 cm thick is required on dielectric cavities and thin spots and/or holes in the plating are common. Thin areas of plating or holes permit electromagnetic coupling between the cavity and the external environment. (2) The low thermal conductivity of suitable dielectrics can result in temperature gradients along the cavity. (3) The low tensile strength of most dielectrics makes the cavity susceptible to breakage during fabrication and processing. (4) The material is expensive and so is the cost of fabrication. (5) The fine tuning must be done by adding on additional coupling loop to the cavity.

The factors relating to the use of metal cavities can also be identified. (1) The plating is not critical and a sufficiently high Q can be achieved without plating. (2) A metal such as aluminum or molybdenum has high thermal conductivity and relieves the problem of temperature gradients. (3) The strength of the metal cavities is very high and breakage due to handling or shock is not a problem. In fact, the metal cavity for a given cavity design lends itself to a light weight structure. (4) The cost of fabrication is considerably cheaper than a corresponding dielectric cavity. (5) Thermal tuning can be used to fine tune the cavity and control the cavity resonance. (6) The relatively high thermal expansion coefficient for aluminum or molybdenum compared to the dielectrics requires a higher temperature stability for the metal cavity structure. It has been demonstrated(13) for an aluminum cavity that the average untuned linear drift rate of the clock, due to the cavity drift, among the

NASA-GSFC NP masers is only 5×10^{-15} per day. Thus for applications of a few weeks or less which is one of the specifications for the GPS system, auto-tuning isn't required for metal cavities except perhaps to help in the initial frequency setting.

By considering all the factors in the cavity design, we decided to make the cavity out of molybdenum and the storage cell from 0.0025 cm thick teflon film. This cavity-cell structure is best suited to meet the stringent requirements placed on a space clock. We have made the cavity using a molybdenum cylinder which is attached to aluminum end plates by heat sweating. The cavity design is 40.6 cm long by 26.7 cm in diameter and the present structure weighs 15 pounds (6.8 kilograms) with significant potential for a lighter cavity. The measured unloaded cavity Q for the structure without plating is 42,000. The measured unloaded cavity Q for a similar all aluminum structure unplated is 50,000. The copper plated molybdenum-aluminum cavity structure has a measured unloaded Q of 60,000. The thin film storage cell can be inserted into the cavity without disassembly of the end plates. (See Figure 14 for photograph of cavity.) The large $\frac{L}{D}$ (length to diameter ratio) of 1.6 is desirable when high performance in stability is the goal. For space applications were size and weight, in addition to the clock performance, are important parameters; a trade-off between clock performance and size and weight will have to be made.

Summary

The mechanical and electrical designs have been completed and components have been tested. The assembly of the clock is now in progress and operation is anticipated during the latter half of 1977. The predicted performance of the clock is a frequency stability of better than 1 part in 10^{14} for averaging times between 50 seconds and 10 days. The main features of the design are: a small fixed quadrupole state selector; a capacitively coupled RF dissociator using an ion-depleted glass bulb; a microwave cavity structure composed of a molybdenum cylinder and aluminum end plates; a cylindrical storage cell made from 0.0025 cm thick FEP "Teflon" film; four magnetic shields; two vacuum chambers; thermal tuning of cavity; and two secondary temperature isolation ovens within the vacuum enclosure. The present engineering model will weigh over 100 pounds but is purposely over designed in weight for the purpose of economy in cost and time and to quickly prove out the performance characteristics of the design. Nevertheless this design, including all of the electronics, can be built to weigh less than 90 pounds and fit into a cylindrical space of 38 cm in diameter by 71 cm long and consume 55 watts of power. Based on knowledge gathered during the design of this unit, future hydrogen clocks will be smaller and lighter than this design and will meet the performance specifications of the GPS system.

Acknowledgments

We are indebted to Dr. H. Weakliem for providing the data on the optical emission spectrum from the dissociator; J. Behrrell for his technical assistance with many of the components; and A. Withey for help with the mechanical design.

References

1. Daniel Kleppner, H. Mark Goldenberg and Norman F. Ramsey, "Theory of the Hydrogen Maser," Phys. Rev. **126**, 603 (1962).

2. H. E. Peters, T. E. McGunigal and E. H. Johnson, "Hydrogen Standard Work at Goddard Space Flight Center," 22nd Freq. Cont. Symp., USAEC, Ft. Monmouth, NJ, 1968; Harry E. Peters, "Characteristics of Advanced Hydrogen Maser Frequency Standards," Goddard Space Flight Center, Preprint X-524-74-15.
3. D. Kleppner, H. C. Berg, S. B. Crampton, N. F. Ramsey, R. F. C. Vessot, H. E. Peters and J. Vanier "Hydrogen-Maser Principles and Techniques," Phys. Rev. **138**, A972 (1965).
4. R. F. C. Vessot, et.al; "Spaceborne Hydrogen Maser Design," Proc. 8th PTI Meeting, NRL, Washington, DC, 1976.
5. W. M. Mueller, J. P. Blackledge, G. G. Libowitz, "Metal Hydrides," (Academic Press 1968). H. H. VanMal, K. H. J. Buschow and A. R. Miedema, "Hydrogen Absorption in LaNi_5 and Related Compounds," J. of Less-Common Metals, **35**, 65 (1974).
6. Allegheny Ludlum Steel Corporation, Lamination Division, River Road, Brackenridge, PA, 15014.
7. E. I. Dupont de Nemours & Co., Plastics Department, Fluorocarbons Div., Wilmington, Delaware, 19898.
8. Victor Reinhardt, "Variable Volume Maser Techniques," Proc. 8th PTI Meeting, NRL, Washington, DC, 1976.
9. Edward M. Mattison, Martin W. Levine and R. F. C. Vessot, "New TE_{111} Mode - Hydrogen Maser," Proc. 8th PTI Meeting, NRL, Washington, DC, 1976.
10. See reference 3. Also, C. Finnie, R. Sydnor and A. Sicard, 25th Annual Symp. on Freq. Contr., Ft. Monmouth, 1971.
11. Harry E. Peters: Goddard Space Flight Center, Tech. Report X-524-74-15 (1973).
12. Stephen F. Jacobs, "Dimensional Stability Measurements of Low Thermal Expansivity Materials using an Iodine Stabilized Laser," Optical Science Center, University of Arizona, Tucson, Arizona 85721.
13. Harry E. Peters: Goddard Space Flight Center, Tech. Report X-524-74-15 (1973).

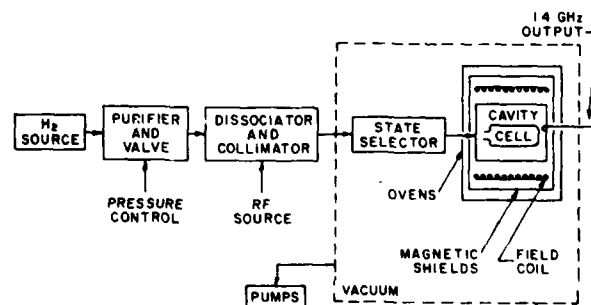


FIG. 1. BLOCK DIAGRAM OF PHYSICS UNIT.

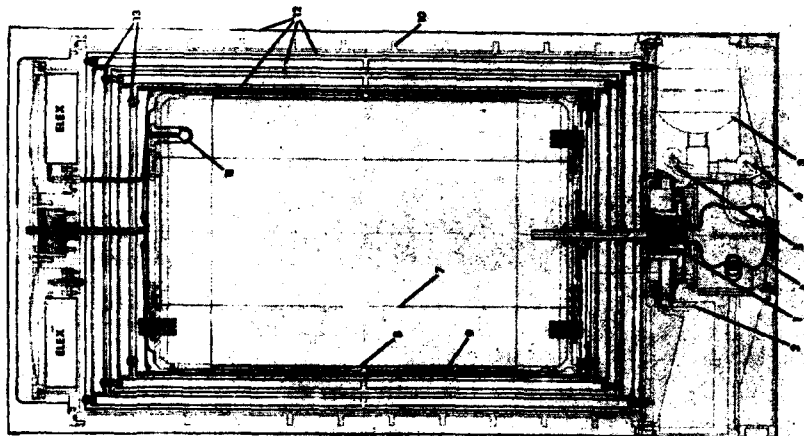


FIG. 2. CROSS-SECTIONAL VIEW: ① STATE SELECTOR ② ION PUMP ③ H_2 STORAGE VESSEL ④ RF DISCHARGE CHAMBER ⑤ PURIFIER ⑥ PRESSURE SENSOR ⑦ TEFLON FILM CELL ⑧ CAVITY STRUCTURE ⑨ FIELD COIL ⑩ VACUUM ENCLOSURE ⑪ COUPLING LOOP ⑫ MAGNETIC SHIELDS ⑬ OVENS.

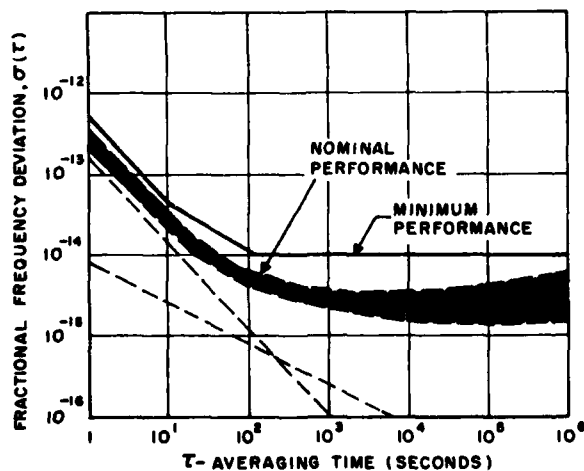


FIG. 3. PREDICTED STABILITY CURVE.

MECHANICAL ASSEMBLY OF HYDROGEN PURIFIER

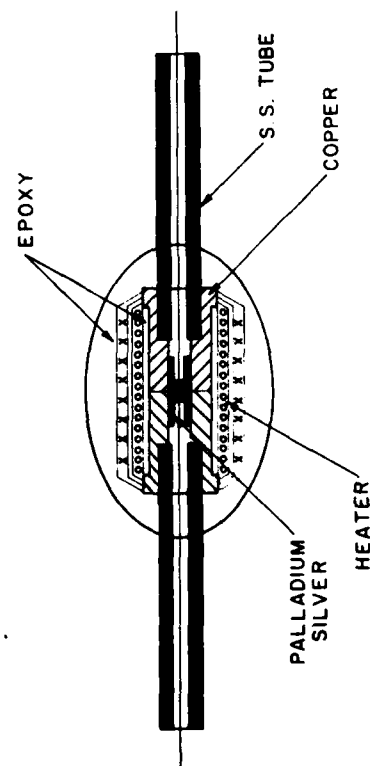


FIG. 4.

PRESSURE IN DISSOCIATOR VERSUS TEMPERATURE OF
PALLADIUM-SILVER PURIFIER NO.2

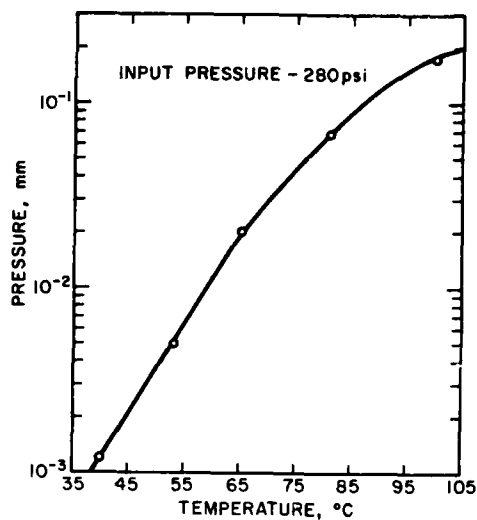


FIG. 5.

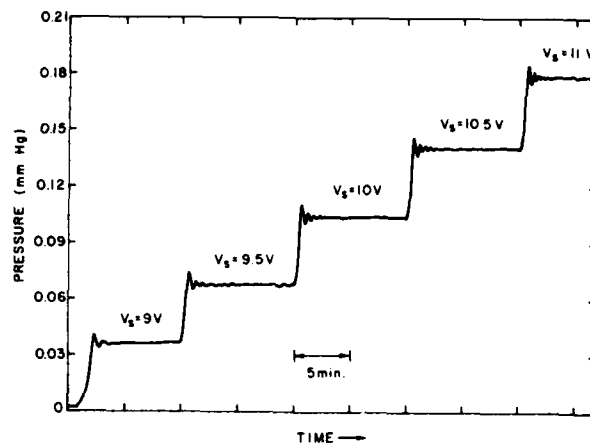


FIG. 6 RESPONSE OF PRESSURE REGULATOR TO STEPS IN
REFERENCE VOLTAGE.

DISSOCIATOR DESIGN

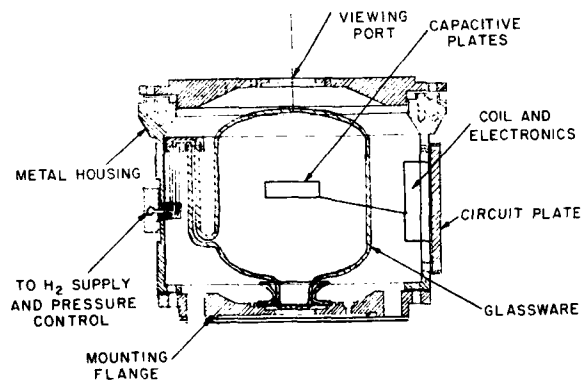


FIG. 7. SKETCH OF DISSOCIATOR ASSEMBLY.



FIG. 8. PHOTOGRAPH OF OPERATING DISSOCIATOR.

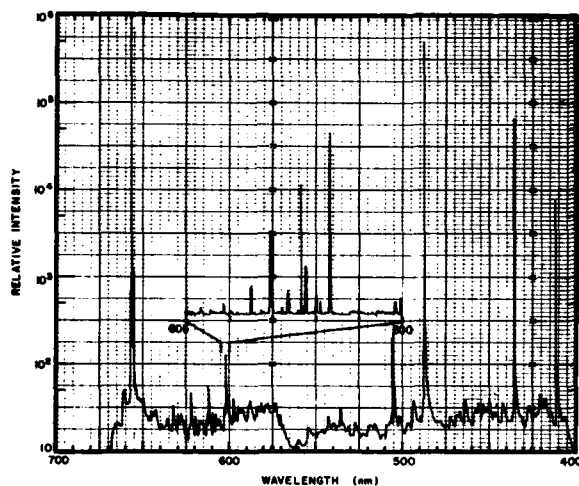


FIG. 9 VISIBLE EMISSION SPECTRUM FROM HYDROGEN DISSOCIATOR (COURTESY OF DR. H. WEAKLIEM, RCA LABORATORIES).

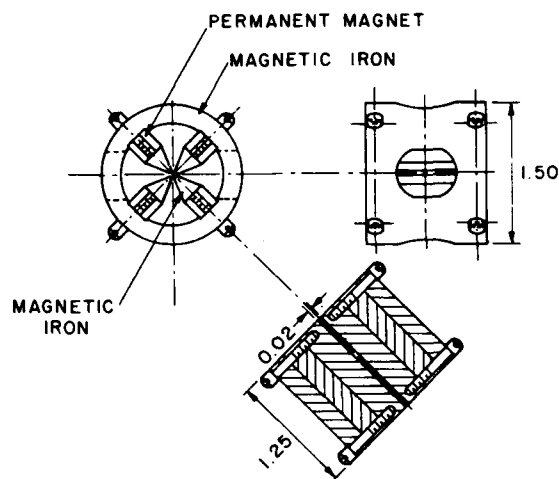


FIG. 10. SKETCH OF STATE SELECTOR ASSEMBLY.

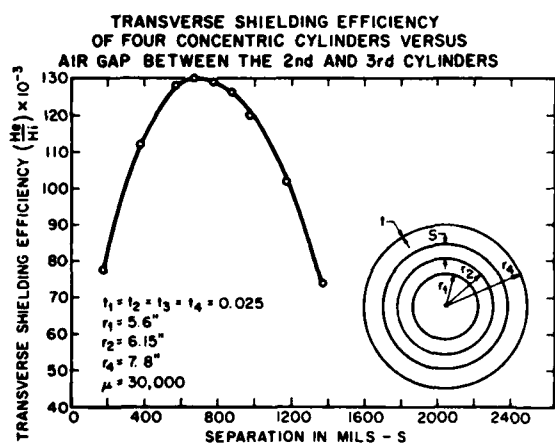


FIG. 11.

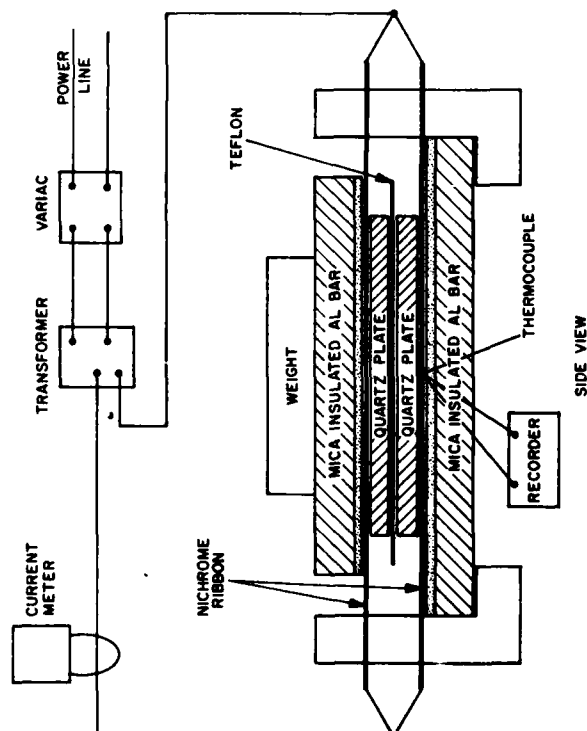


FIG. 12. SKETCH OF APPARATUS FOR HEAT SEALING OF TEFLON FILM.

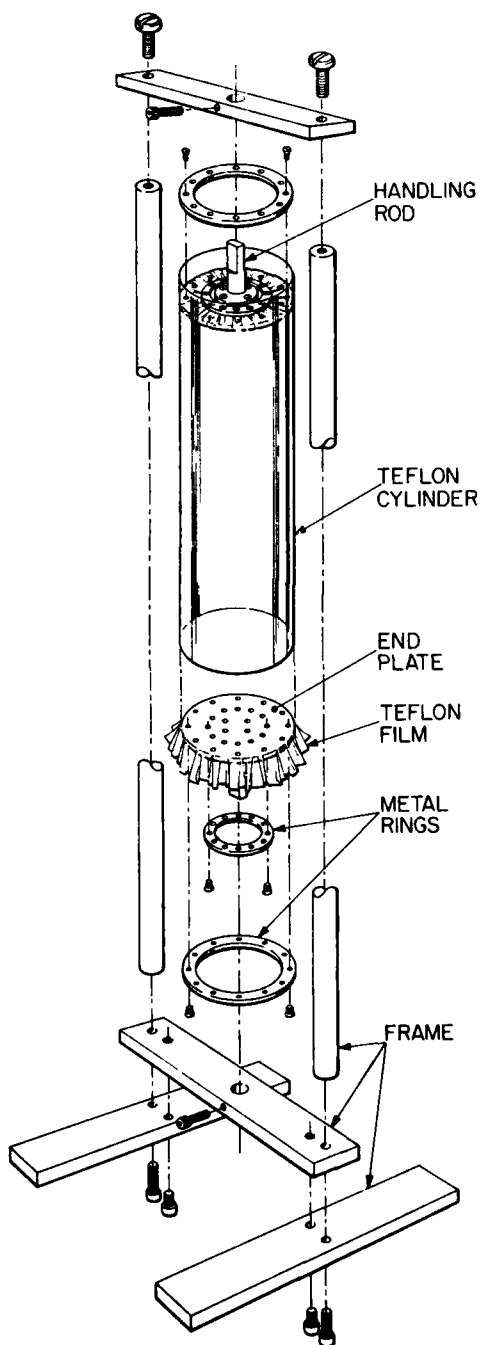


FIG. 13. ASSEMBLY OF TEFLON FILM CELL.

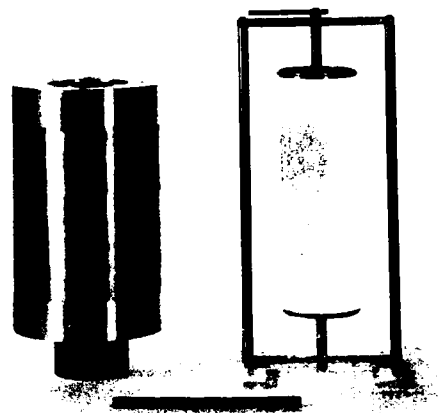


FIG. 14. PHOTOGRAPH OF CAVITY AND TEFLON FILM STORAGE CELL.

TABLE I. CAVITY AND STORAGE CELL PROPERTIES

COMPONENT	THERMAL COEFFICIENT OF EXPANSION	TEMPERATURE DEPENDENCE OF SHIFT IN CAVITY FREQUENCY	TEMPERATURE STABILITY, ΔT , REQUIRED FOR FREQUENCY STABILITY $\times 10^4$ OF 1 PART IN 10^4
CER-VIT CAVITY	$1.5 \times 10^{-7}/^\circ\text{C}$	180 Hz/°C	7.9×10^{-3}
ALUMINUM CAVITY	$25 \times 10^{-6}/^\circ\text{C}$	29,900 Hz/°C	4.7×10^{-5}
MOLYBDENUM CAVITY	$5 \times 10^{-6}/^\circ\text{C}$	5,980 Hz/°C	2.4×10^{-4}
GFEC CAVITY	DIA: $1 \times 10^{-6}/^\circ\text{C}$ LONG: $1 \times 10^{-6}/^\circ\text{C}$	1200 Hz/°C	1.2×10^{-3}
FUSED QUARTZ CELL 1 MM THICK	$0.5 \times 10^{-4}/^\circ\text{C}$ ⁽¹⁾	400 Hz/°C	3.5×10^{-3}
TEFLON FILM CELL 0.001 INCH THICK	$6.4 \times 10^{-4}/^\circ\text{C}$ ⁽¹⁾	48 Hz/°C	2.9×10^{-2}

(1) Temperature dependence of dielectric constant.

(2) Assumed $C_0/C_1 = 1 \times 10^{-5}$.

HYDROGEN MASER WITH A DOUBLE CONFIGURATION BULB FOR WALL SHIFT MEASUREMENTS IN THE TEMPERATURE RANGE 25-120°C

P. Petit, M. Desaintfuscien and C. Audoin
Laboratoire de l'Horloge Atomique
Equipe de Recherche du CNRS
Bâtiment 221 - Université Paris-Sud
91405 Orsay - France

Summary

Possible ways to improve the H-maser accuracy are stated. Among them, the use of a H-maser of regular design, equipped with a special bulb, made of quartz and showing two coaxial parts is described. A small quartz valve allows to use the solid bulb in two different configurations, with two different mean diameters. This device is operated in the temperature range 25-120°C, and preliminary results are reported. A change of the temperature dependence of the wall frequency shift, for a FEP 120 coating, is observed at 360 K and the wall shift goes to zero at 380 K about.

Introduction

In all presently operated atomic frequency standards, it is not possible to interrogate the atoms in ideal conditions, i.e. without interaction with any field, and at rest. In the hydrogen maser, the most annoying spurious effect is related to collisions of hydrogen atoms with the wall of the containing vessel. If $\bar{\phi}$ is the mean phase shift per collision, the macroscopically observed frequency shift Δf_w of the hyperfine transition of stored hydrogen atoms is

$$\Delta f_w = \frac{n}{2\pi} \bar{\phi} \quad (1)$$

where n is the mean collisions rate of hydrogen atoms with the wall. In general, n is related to the area to volume ratio of the bulb, A/V , and we have :

$$\Delta f_w = \frac{\bar{v}}{8\pi} \frac{A}{V} \bar{\phi} \quad (2)$$

where \bar{v} is the mean atomic velocity in the bulb. For a spherical bulb, $A/V = 6/D$ where D is the diameter, and then

$$\Delta f_w = \frac{3\bar{v}\bar{\phi}}{4\pi D} \quad (3)$$

It is well known that, presently, the properties of successive wall coatings suffer from a lack of reproducibility, as far as the wall shift Δf_w is concerned. This affects the precision in the determination of the wall shift, by the classical method which consists in substituting spherical coated bulbs of different diameters, in a maser¹. This is the present limiting factor of the H-maser accuracy, to the 10^{-12} level, whereas the contribution of the other sources of frequency offset, to the error budget, is smaller than 1×10^{-13} .

This work has been sponsored by Centre National de la Recherche Scientifique and Bureau National de Métrologie.

The aim of the experiment described here is to investigate a possible mean to improve the accuracy of a H-maser frequency standard of classical design. We report preliminary results achieved with a special bulb in the temperature range 25-120°C, which demonstrate the feasibility to operate a H-maser frequency standard at the temperature which cancels the wall shift of presently used bulb coatings.

Possible solutions of the wall shift problem

Several solutions of this problem have been proposed :

1) the first attempt to overcome the wall shift limitation has been the big box maser of Harvard University³ in which a bulb showing a mean diameter of about 1.5 m was used in a maser of special configuration. The mean collision rate n , was then 10 times smaller than in masers of regular design^{4,5}.

2) the use of a bulb, a part of which is made of a thin membrane^{6,7}, allows to vary the volume of the bulb, as sketched on figure 1b. For the two extreme positions of the flexible membrane, one then simulates two bulbs of different mean diameters, but operating with the same surface (however, stresses are developed in the bended region, which are different for the two positions, and might change the mean phase shift per collision, on that part of the surface). The application of this idea to the big box maser (fig. 1a) allowed an experiment in which the wall shift contribution to the error budget was 2.6×10^{-13} only⁸.

The implementation of an elongated cylindrical bellows⁹ (fig. 1c) rely on the same idea. It offers the possibility of a measure of the wall shift as a continuous function of the surface to volume ratio of the bulb. Furthermore, it is expected that possible stress effects would cancel on regions of opposite curvature of the convolutions.

The bulb shown on figure 1d has also been considered¹⁰. The piston allows a continuous variation of the ratio A/V .

These methods where a flexible membrane or a piston is used have the common drawback to cause difficulties as far as the mechanical stability of the storage bottle, the temperature coefficient of the cavity and the recovery of large cavity detunings are concerned. Furthermore, in an oscillating maser, the change in the area to volume ratio is limited by the requirement to give the product of the filling factor^{4,3} times the bulb volume a value compatible with the oscillation condition.

3) one of us (P.P.) proposed to remove these last drawbacks by using a solid bulb, made of quartz, designed with two coaxial parts and a small valve as shown on figure 1e. For each of the two stable positions of the valve, the area to volume ratio of the region occupied by atoms can differ by a large factor (it is

even possible, in principle, to add quartz walls, in the outer part of the bulb, increasing the area without an appreciable change of the volume). The A/V ratio variation can then be much larger than allowed with a flexible bulb.

According to the valve position, the surface interacting with hydrogen atoms are different. But, the quartz surfaces being lined in the same coating operation, a good enough homogeneity of the wall properties can be expected. This point can be checked by operating several bulbs of this kind.

The bulb shown on figure 1a benefits of a better centering in the microwave and the static magnetic fields, which is of interest to minimize the frequency shift which might be connected to the inhomogeneities of these fields ¹¹. This is the bulb used in the reported experiment.

The idea of the small valve allowing to change the configuration of the storage region can also be applied to a bulb with a septum orthogonal to its axis of symmetry ¹².

4) Presently used bulb coatings show a null of the mean phase shift per collision at a given temperature close to 100°C ^{13,14,15}. At that temperature, the transition frequency and thus the oscillating frequency is not shifted as an effect of wall collisions, and ¹⁶ this can be experienced by modulating the A/V ratio by one of the means described above. Furthermore, at that point, the exact knowledge of the involved A/V ratios is not needed. These ratios only determine the sensitivity of the method.

5) One obvious way, but still speculative, is to search for coatings showing reproducible enough properties (in that event, it would not be needed that they show a null of the mean phase shift per collision, although an useful characteristics when occurring at a convenient temperature). In that respect, it would be interesting to investigate more the properties of some fluoropolymers. For instance, FEP 120 behaves as a perfect solid for $T < 200$ K and its properties might be much more reproducible there than for higher temperatures ¹⁷. The change of properties at $T \approx 200$ K is due to a second order phase transition in the bulb material. This temperature might be increased to a more tractable value by reticulating the polymer chains (without the introduction of paramagnetic centers).

Experimental set-up

A hydrogen maser of classical design ¹⁸ has been equipped with a special bulb and it has been slightly modified in order to allow operation in the temperature range 25-120°C. Figure 2 shows details of the bulb used and of its setting in the microwave cavity.

Double configuration bulb

It is made of quartz and is FEP 120 coated. It comprises two coaxial parts, with a common bottom. It is firmly held in position with the help of a cone-shaped grinding, in its lower part, and of a spring-grounded quartz tube in its upper part. A small valve, made of quartz (diameter : 24 mm, thickness : 4 mm) and FEP 120 coated, showing two spherical grindings, rests on one of the two spherical-grinded seats. A quartz rod is attached to the valve and is driven through amagnetic bellows. All gaskets are still metallic ones in the modified H-maser.

The valve allows to use the solid bulb in two different configurations, according to its position. In the configuration n°1, the valve sits on the top of the inner part of the bulb. The region offered to hydrogen atoms is then a bulb with an equivalent mean diameter of 15.3 cm about. In the other stable configuration, n°2, the valve closes the outer part of the bulb

and allows a communication between the inner region and another region which is coaxial with the previous one. The surface where the atoms collide is then multiplied by a factor larger than three, whereas the volume is not too much increased. The equivalent mean diameter of the bulb in configuration N°2 is 8.3 cm about (as stated above, it could be made smaller by inserting coated quartz walls in the interval between the limits of the two useful regions). The main characteristics of this bulb are summarized in the following table

Configuration	$V(\text{cm}^3)$	$A(\text{cm}^2)$	$\bar{D}(\text{cm})$	$\frac{V_b}{\eta V_c}$	$T_b(\text{s})$
n° 1	2483 ±1	973 ±20	15.3	0.38	0.75
n° 2	3990 ±1	2685 ±60	8.3	0.38	1.21

where \bar{D} is the equivalent mean diameter, η is the filling factor, V_b the total volume offered to atoms, V_c is the volume of the cavity and T_b is the calculated geometrical life time at a temperature of 20°C. The communication time constant between the two parts of the bulb is 8 ms about. It is small enough compared to T_b so that the atomic density can be considered as uniform in configuration n° 2. Similarly all the H-maser properties based on motional averaging ^{4,11} are still valid in configuration n° 2. In this configuration, the H-maser behaves -except for wall interactions- as if the septum would be removed. Due to the valve smallness and to the fact that it moves in a region of weak electric field, the cavity resonant frequency differs by 12 kHz only for the two valve positions.

Peculiar attention has been paid to ensure homogeneous properties of the FEP 120 coating. The main features of the procedure are i) slow and smooth rotation of the bulb containing a small amount of FEP 120 suspension, ii) dripping in the presence of a humid atmosphere, iii) rotation during baking up to 360°C, in the presence of an oxygen flow in the two parts of the bulb and iiiii) slow cooling in order to obtain a similar texture for all the coated surfaces. The bulb received two layers of coating, as well as the valve and the upper quartz tube which grounds the bulb.

Microwave cavity and thermal control

The temperature coefficient of the microwave cavity has been minimized, in view of the large temperature range to be explored in preliminary experiments.

The cavity structure is made of quartz, as shown on figure 2, so that all parts in contact have the same thermal coefficients. Jumps in the resonant frequency are then avoided. The thermal coefficient of a cavity, made of quartz walls, without bulb, is 0.6 kHz.K^{-1} about. In the presence of the double configuration storage bulb, it rises up to 3.5 kHz.K^{-1} . A thermal compensation is then required. It is insured by three posts in the upper part of the cavity. They comprise quartz and copper rods of adjustable length. It is then possible to give these posts the desired thermal expansion coefficient. One proceeds by trial and error to minimize the thermal coefficient of the microwave cavity, in the presence of the storage bulb. This is made necessary by the large uncertainty in the value of the thermal coefficient of the dielectric constant of quartz. Finally, a mean thermal coefficient of 0.1 kHz.K^{-1} is achieved in the temperature range 25-120°C. A plunger, with a tuning range of 30 kHz allows to approach the cavity tuning position in the whole temperature range, whatever the valve position is. A varactor allows the fine tuning of the cavity.

The microwave cavity is enclosed in a thermal shield, made of copper and situated within the vacuum tank. It ensures the thermal homogeneity which is required for a proper operation of the thermal compensation set. Furthermore, it improves the decoupling between the microwave cavity and the remaining parts of the maser.

The driving mechanism of the valve is thermally controlled. For each temperature of the bulb, the temperature of most of the different controlled walls is properly adjusted, by changing the reference resistors of the temperature sensing bridges¹⁹. The reference resistors and the electronic circuits (except the power stage) ensuring the thermal control of the inner parts of the maser are also temperature regulated.

Experimental results

Frequency stability

When the modified maser is operated with a temperature of the bulb and the microwave cavity in the range 26-104°C, its measured frequency stability looks the same as it was before modification¹⁷. One obtains (in a bandwidth of 3 Hz)

τ (s)	1	10	10^2	10^3	10^4
$\sigma_y(\tau)$	4×10^{-13}	4×10^{-14}	8×10^{-15}	4×10^{-15}	2×10^{-14}

for each position of the valve. For $\tau > 10^3$ s, the measured frequency stability might be limited by the reference maser in which the microwave cavity still shows a thermal coefficient of 1.5 kHz.K⁻¹. It was noticed a slight decrease of the frequency stability at a temperature of 122°C. However, in this measure, the steady state conditions were not completely reached.

These results show that the mechanical stability of the device is quite good.

Temperature variations of the wall shift

After the temperature reached the desired value within 0.03K about, and the background pressure in the bulb compartment lowered to less than 10^{-5} Pa (which requires a delay of approximately 12 days between each temperature setting), the modified maser is spin-exchange frequency tuned¹¹. As the temperature rises, the contribution of the wall collisions to the relaxation times T_1 and T_2 and the contribution of H-atoms recombination to the storage time constant T_b increases¹⁵. The intensity of the hydrogen beam must be increased accordingly. It results that the range of beam intensities which is allowed for cavity tuning is narrowed. However, even at the temperature of 122°C, the cavity can be tuned such that the error in the oscillating frequency is smaller than 0.1 mHz.

The frequency is measured, for both valve positions, relative to the frequency of a reference maser. Corrections are then applied to account for the second order Doppler effect, and the second order magnetic field frequency shift. The value of the static magnetic field is close to 8 mG and it has been carefully checked, for both valve positions that the magnetic field inhomogeneities¹¹ do not shift the observed frequency.

Figure 3 shows the preliminary results which have been obtained. One sees that the temperature dependence of the wall shift follows two different laws, according to the temperature range. This corresponds, likely, to another second order phase transition occurring at a temperature close to 380 K, which was not reported previously in H-maser works. The null of the

wall shift is reached at a temperature close to 380 K.

Future experiments will concentrate measurements on the temperature range 360-395 K.

Conclusion

It has been experimentally proved that a hydrogen maser, equipped with a double configuration storage bulb, can conveniently be operated as a frequency standard at a temperature close to 100°C, where the wall shift goes to zero. This allows to investigate one way to improve the H-maser accuracy.

References

1. J. Vanier, R. Larouche and C. Audoin, Proceedings of the 29th Annual Symposium on Frequency Control p. 371 (1975) copies available from E.I.A.
2. P.W. Zitzewitz, E.E. Uzgiris and N.F. Ramsey, Rev. Sc. Instr. 41 81 (1970)
3. E.E. Uzgiris and N.F. Ramsey, Phys. Rev. A1, 429 (1970)
4. H.M. Goldenberg, D. Kleppner and N.F. Ramsey, Phys. Rev. 123, 530 (1961)
5. D. Kleppner, H.C. Berg, S.B. Crampton, N.F. Ramsey, R.F.C. Vessot, H.E. Peters and J. Vanier, Phys. Rev. A 138, 972 (1965)
6. D. Brenner, Appl. Phys. 41 2942 (1970)
7. P.E. Debely, Proc. of the 24th Annual Symposium on Frequency Control, p. 259 (1970). Copies available from E.I.A.
8. V.S. Reinhardt and J. Lavanceau, Proc. of the 28th Annual Symposium on Frequency Control, p. 379 (1974)
9. H.E. Peters, Proc. of the 29th Annual Symposium on Frequency Control p. 362 (1975) copies available from E.I.A.
10. J. Vanier, private communication
11. S.B. Crampton and H.T.M. Wang, Phys. Rev. A 12 1305 (1975)
12. J. Vanier, private communication
13. R.F.C. Vessot and M.W. Levine, Metrologia 6 116 (1970)
14. P.W. Zitzewitz and N.F. Ramsey, Phys. Rev. A 13 51 (1971)
15. M. Desaintfuscien, J. Viennet and C. Audoin, Proc. Second Seminar on Frequency Metrology - Copper Mountain - USA - p. 565 (1976)
16. R.F.C. Vessot, M.W. Levine, P.W. Zitzewitz, P. Debely and N.F. Ramsey, Proc. of the International Conference on Precision Measurement and Fundamental Constants, Gaithersburg - USA - p. 27 (1970)
17. M. Desaintfuscien, Private communication
18. P. Petit, J. Viennet, R. Barillet, M. Desaintfuscien and C. Audoin, Metrologia 10 61 (1974)
19. R. Barillet, J. Viennet, P. Petit and C. Audoin, J. of Physics E. Scientific Instruments 8 544 (1975)

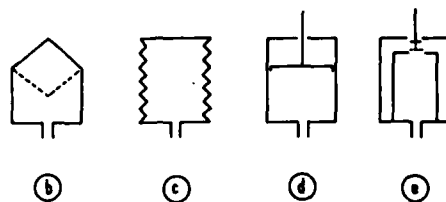
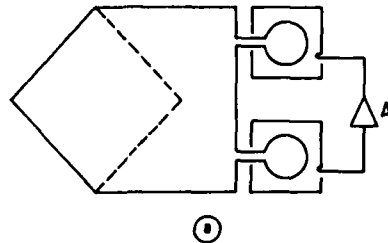


Fig. 1

- 1 - Various thoughts concerning the use of a bulb for wall shift measurements.
- The Harvard University big box maser with a flexible bulb ^{2,8} (The size of the bulb is ten times as large as in masers of regular design)
 - Bulb with a flexible membrane ^{6,7}
 - Bulb as an elongated cylindrical bellows, made of teflon ⁹
 - Solid bulb with a piston ¹⁰
 - Solid bulb with two coaxial parts and a valve allowing the operation with two different mean diameters. The two parts are best centered in the microwave cavity and in the static field.

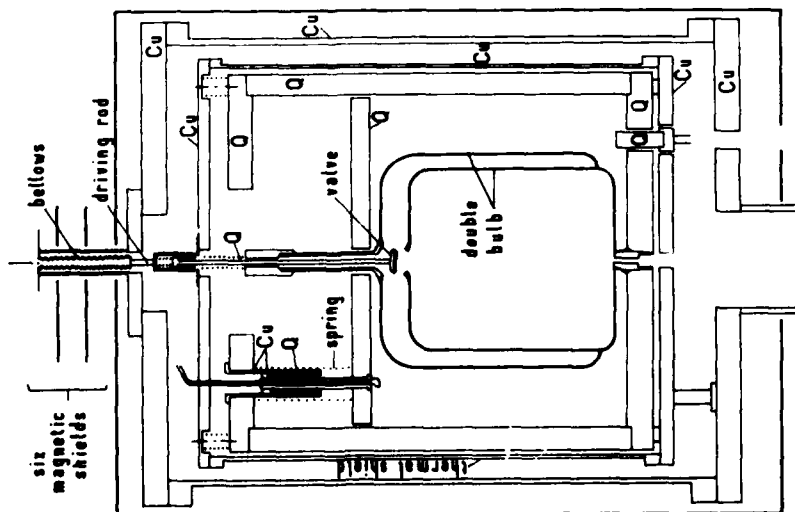


Fig. 2

2 - The double configuration storage bulb and the thermally compensated microwave cavity. Q and Cu mean quartz and copper, respectively.

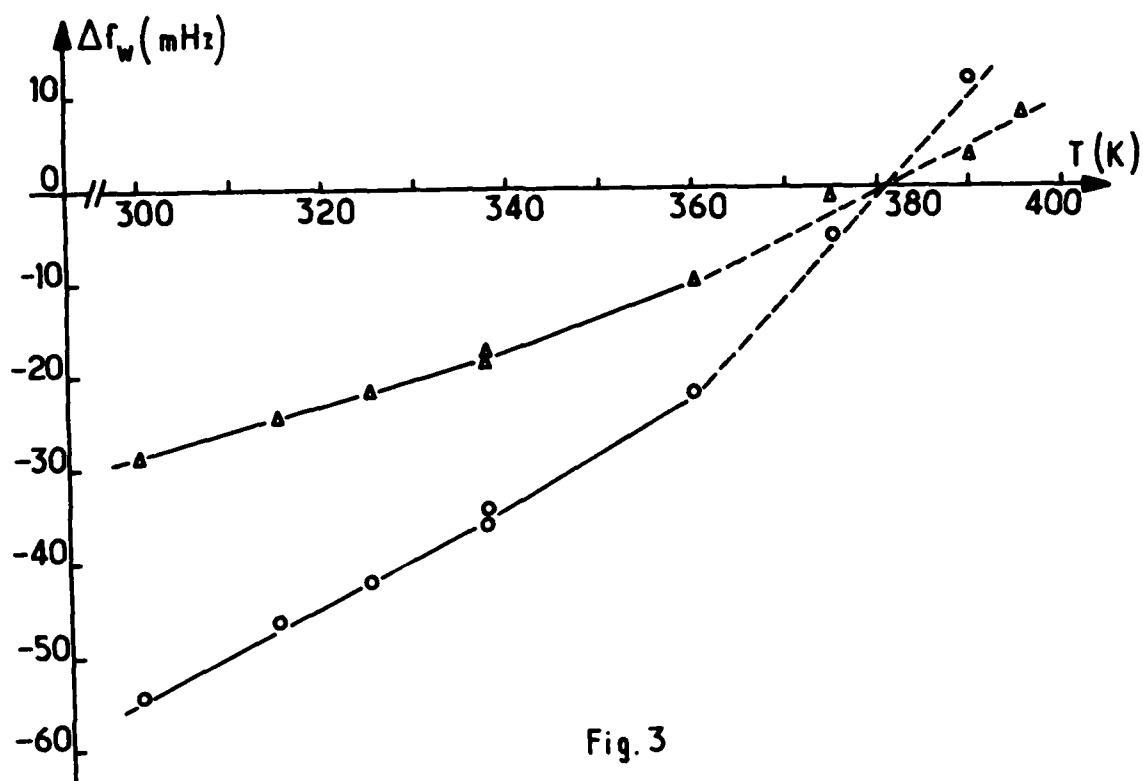


Fig. 3

3 - Temperature dependence of the wall shift, for a FEP 120 coating. The triangles represent experimental points obtained with the valve down (configuration n° 1) and the dots represent experimental points obtained with the valve up (configuration n° 2). The origin of the wall frequency shifts has been adjusted at the curve crossing.

A NEW GENERATION OF SAO HYDROGEN MASERS

M. W. Levine, R. Vessot, E. Mattison, G. Nystrom, T. Hoffman, and E. Blomberg
Smithsonian Astrophysical Observatory
Cambridge, Massachusetts

Summary

A new generation of hydrogen masers for field use has been developed at the Smithsonian Astrophysical Observatory. The VLG-11 series incorporates a number of improvements over the earlier VLG-10 masers, including upgraded thermal performance, freedom from barometric-pressure influences, and thoroughly modernized and field-repairable electronics.

This paper describes the construction of the VLG-11 "physics package," the supporting electronics system (thermal, pressure, magnetic field), and the signal-processing system (phaselock receiver and digital synthesizer). The features of the physics package are a new, lightweight storage-bulb design, a rugged Cervit cavity carefully isolated from pressure variations on the vacuum tank, a multilayer printed-circuit solenoid, which provides an exceptionally uniform "C" field, and torrispherically domed magnetic shields supported by machined semirigid foam. The thermal control system provides two-level, five-zone temperature regulation to minimize environmentally induced temperature gradients at the maser cavity. The cooling system for the hydrogen dissociator is a temperature-controlled closed-circuit air-to-air heat exchanger that further reduces the influence of ambient-temperature variations on the cavity resonant frequency. The digital synthesizer was specifically designed for this application, providing a 0.001-Hz step resolution over a tuning range of 405,750.00 to 405,753.999 Hz. The synthesizer combines very low power consumption, less than 3 watts, with the capability of operating directly from 28-volt battery back-up power sources.

Performance data on the VLG-11 hydrogen masers — including short-term frequency stability, frequency vs. ambient temperature, frequency vs. barometric pressure, and frequency as a function of ambient magnetic field — are presented. Long-term frequency stability data are also covered, to the extent that such data are available.

Key words (for information retrieval). Hydrogen Maser, Frequency Standard, Frequency Stability, Oscillator, Design.

Introduction

The design of the Smithsonian Astrophysical Observatory (SAO) masers is the result of an unbroken evolutionary process that began in 1960, shortly after the invention of the maser by Kleppner, Goldenberg, and Ramsey.¹ In 1962, the hydrogen maser group at Varian Associates, Beverly, Massachusetts, began a program for the development of a transportable, self-contained hydrogen maser under the sponsorship of the National Aeronautics and Space Administration/Marshall Space Flight Center. Shortly thereafter, a spacecraft design effort was begun, with the goal of measuring the gravitational redshift using a payload launched into a 24-hour elliptical orbit by a Titan 3C booster.² Although this approach for redshift measurements was abandoned because of the high cost of an orbital satellite program, the design later evolved into the VLG-10A ground maser.

The maser program, which began at Varian Associates and was continued by Hewlett-Packard when it acquired the Varian frequency and time operation, was transplanted to SAO in Cambridge, Massachusetts, in 1969. SAO built a series of VLG-10A masers for the Naval Research Laboratory and for a consortium of four radio-astronomy observatories. In 1970, the gravitational redshift program was revived, and three additional VLG-10A masers were built to serve as the ground reference standards for the redshift experiment.³

The requirements for the redshift probe maser proved to be extremely demanding; the clock was to be launched in a 200-lb payload by a solid-fuel Scout rocket to an altitude of 10,000 km, staying aloft for 2 to 3 hours. The probe maser was required to survive launch environment and operation within specifications immediately after burnout — necessitating the design of an extremely rugged, lightweight cavity structure and a sophisticated thermal control system. The qualification model of the probe maser was subjected to intensive shock, acoustic, vibration, spin, magnetic-field, and thermal-vacuum testing for 9 months.

The VLG-11 is the end result of two converging lines of development — the earlier ground masers and the innovative probe maser — and incorporates the best features of both designs into an entirely new generation of hydrogen masers.

VLG-11 Design Philosophy

There are two basic aspects to the design philosophy developed at SAO for the VLG-11 series of hydrogen masers. First, the maser cavity was to be as inherently stable as possible. Then a very tightly controlled thermal and mechanical environment was to be established for the cavity to permit maser operation over long periods of time without autotuning or other adjustment. Toward this end, the cavity was constructed of Cervit 101, a material with a very small thermal-expansion coefficient (2 to $5 \times 10^{-7}/^{\circ}\text{C}$) and long-term resistance to dimensional creep. To reduce the aspect of the thermal coefficient of dielectric loading, the mass of the storage bulb was reduced to 175 g (from 350 g as used in the VLG-10 masers) without sacrificing the stiffness-to-weight ratio.

The second basic consideration was that the VLG-11 was to be a readily transportable, rugged, and completely integrated frequency standard. The entire system was engineered as a unit, and all electronics, including the signal-processing elements (receiver and synthesizer), were mechanically and electrically integrated into the overall design. The end result is a relatively small instrument weighing about 650 lb and requiring a source of only 24 to 30 volts DC at 200 watts to operate.

Mechanical Design

The maser frequency standard is housed in a standard, caster-equipped, aluminum cabinet measuring 52 inches high by 22 inches wide by 30 inches deep. Three aluminum brackets are welded directly to the cabinet frame to engage a set of ears on the main support plate of the cylindrical physics package. The electronics

packages are built up around the physics packages on the front and rear mounting rails of the cabinet.

The physics package is a completely self-contained unit, interconnected to its support systems through a single 61-pin cylindrical connector for power, RF output, control and monitoring lines, and two hoses for recirculating the air used to control the RF dissociator temperature.

The electronics package includes the control assembly, consisting of a monitor panel, main card file, control panel, receiver box, synthesizer box, and distribution unit. Figure 1 shows the physical placement of these components. On the rear panel are the air-flow control assembly, the heat-sink assembly for the power transistors, and the high-voltage power supply for the ion pump.

In Figure 2, the control assembly is swung out on its hinges with the distribution unit partially withdrawn from the cabinet on its chassis slides. The hinging arrangement permits convenient access to the rear of the card file, monitor panel, and control panel for maintenance or service without interruption of power. The chassis slide provides similar capabilities for the distribution unit, which houses all low-voltage power supplies and analog monitoring buffer amplifiers. Figure 3 shows the front of the card file with the individually shielded printed-circuit cards plugged into the backplane.

A closeup of the receiver and synthesizer boxes with the front covers removed is seen in Figure 4. All the synthesizer modules and most of the receiver modules are individually shielded plug-in boxes, which can be easily removed from the front for service.

Physics Package

The figures in this section illustrate the design evolution from the VLG-10A series of masers to the new VLG-11. The major changes are a more rugged cavity structure, adapted from the probe maser; improved thermal design, also an outgrowth of the probe maser program; and improved magnetic shielding through the use of torrispherical shields.

Figure 5 illustrates the cavity-bulb assembly, including the output line and isolator, which is properly part of the cavity resonator. The photograph of the lightweight bulb in Figure 6 reveals the tetrahedral strut supports. Figure 7 shows the cavity-support structure and outlines the major load paths. Stress changes on the cavity owing to the expansion coefficient of the aluminum hold-down cylinder are absorbed by the Belleville washer. Similarly, thermally induced radial stress at the base is relieved by the rollers. The cavity is isolated by barometrically induced stress provided by the "floating base" construction shown in the figure.

The magnetic shield-oven assembly surrounding the vacuum tank is shown in Figure 8. The torrispherical shields used in the VLG-11 are supported at their ends by rigid foam-glass insulation. These shields are subjected to far less local stress than the flat-ended shields in the VLG-10A and have improved lapped joints at the lower cover (the upper cover is not removable). The printed-circuit solenoid for the VLG-11 is similar in design to the probe solenoid, though considerably larger in size. The multilayer construction allows nearly complete cancellation of stray magnetic fields and provides a rugged structure that fits closely within the innermost magnetic shield surrounding the vacuum tank. Figure 9 is a photograph

of the printed-circuit solenoid.

Figure 10 illustrates the thermal control system of the physics package, showing the division of the vacuum tank into three independent zones and the oven into two zones. The thermal control system is described in detail in the following section.

The ion-pump manifold, dissociator, and state selector for the VLG-10 and VLG-11 masers are compared in Figure 11. The dissociator glassware in both cases is similar to the probe maser design. The location of the RF discharge optical monitor is shown in the figure; the operation of this circuit is described in the next section.

Electronic Systems

There are two basic categories of maser electronic systems: 1) maser support (thermal, magnetic, pressure, and tuning control) and 2) maser signal processing. Unlike earlier SAO designs, the VLG-11 has all the electronics integrated into a single functional unit with all the controls, metering, and power supplies centralized. The support and processing electronics are mechanically and electrically compatible with each other and with the physics package.

Maser Thermal Controls

Experience with earlier models of hydrogen masers demonstrated that the sensitivity of the cavity tuning to ambient temperature changes was much greater than expected. Increasing the thermal gain of the oven controllers and improving the temperature stability of the thermistor bridge preamplifiers did not produce commensurate improvements in temperature coefficients. Computer modeling of the maser system for the probe maser design indicated that thermal gradients along the vacuum tank are the primary contributor to the ambient-temperature sensitivity of the cavity. Exhaustive thermal testing of the probe maser confirmed the validity of the model. Increased thermal gain or improved amplifier stability is ineffectual in attacking this problem; however, division of the tank (and oven) surfaces into independently sensed and controlled zones has been shown to be a powerful technique in minimizing gradients.

The vacuum tank, therefore, is divided into three zones - dome, cylinder, and base - each with its own sensing thermistor, amplifiers, and heater windings. Each zone can respond independently to external thermal loading without materially affecting the temperature of the other zones. To minimize the thermal stress on the vacuum-tank controllers, the oven is divided into a dome-cylinder zone and a base zone, each of which is independently controlled. The major heat-flow path into the oven is by way of the base zone, which is adjacent to the pump and thermally coupled to it by the neck structure and mounting studs. Independent control of the base zone permits it to respond to its varying loads without excessive heating or cooling of the relatively well-insulated cylinder dome zone.

A separate oven is provided for the isolator and the first RF preamplifier. The isolator box is mounted on the upper electronics assembly, along with four printed-circuit boards - three tank preamplifiers and the tuning-diode voltage regulator. The upper electronics assembly, in turn, is temperature-controlled by the air-flow system, which also provides controlled-temperature air to the dissociator.

The close-cycle air-flow system is a major departure from previous practice. Air is drawn from a

plenum chamber in the air-flow assembly, heated to approximately 38°C by an actively controlled heater, and circulated by means of two blowers through separate circuits to the dissociator assembly and to the upper electronics/ion-pump assembly. The air flow is then returned to the plenum, passing through an air-to-air heat exchanger, one side of which is in equilibrium with ambient room air.

The design goal for both the oven and the isolator box is a temperature stability of $\pm 0.01^\circ\text{C}$ over a 20 to 30°C range of ambient temperatures. The goal for vacuum-tank stability under the same conditions is $\pm 0.0005^\circ\text{C}$, and that for the temperature-controlled air in the air-flow system, $\pm 0.5^\circ\text{C}$.

Preliminary measurements indicate that the performance requirements on the two oven zones have been met. The resolution of the temperature-monitoring equipment is approximately $\pm 0.001^\circ\text{C}$, so full verification of vacuum tank performance is not possible; however, the tank zones are controlled within the resolution limits of the instrumentation. The air-flow system has not been fully tested at this time.

The six thermal control amplifiers are printed-circuit boards that plug into the main card file in the control assembly (see Figure 3). The three most critical preamplifier modules, for the three tank zones, are mounted in the temperature-controlled upper electronics assembly. The preamplifiers for the two oven zones and the isolator box are mounted on the amplifier boards in the control assembly. The power transistors for the six heaters are mounted external to the maser cabinet on the rear-mounted heat-sink assembly. The controlled-temperature air system is completely self-contained, with integral electronics, in the rear mounted air-flow assembly.

Hydrogen-Pressure Control

The hydrogen supply for the dissociator is similar to that used in earlier SAO ground masers. A 2- to 3-year supply of hydrogen gas is stored in a 1-liter tank at about 1000 psi. A thermally controlled palladium valve meters the hydrogen into the dissociator. Hydrogen pressure is sensed at the dissociator input by a thermistor Pirani gauge. A second identical thermistor senses the high-vacuum side of the pumping system, providing ambient-temperature corrections to the pressure-measurement circuitry. The output of the thermistor bridge drives amplifiers that control the temperature of the palladium valve so as to maintain the hydrogen pressure at a constant level.

The pressure-control electronics consist of a printed-circuit card that is installed in the main card file. The power transistor that drives the palladium-valve heater is mounted on the heat-sink assembly on the rear of the maser cabinet.

Dissociator Oscillator

The RF oscillator in the VLG-11 dissociator is a refinement of the one-transistor design used in earlier SAO masers. The oscillator operates at approximately 30 MHz with 6 to 10 watts of DC power input.

The VLG-11 version of the dissociator oscillator is built up on a printed-circuit board and employs microstrip components and techniques to minimize stray inductance and capacitance. The result is a very reliable, stable, and reproducible circuit. A four-turn coil serves as both the oscillator tank inductor and the electrode for excitation of the atomic hydrogen discharge.

The oscillator supply voltage is controlled by voltage-regulator circuitry on a printed-circuit card mounted in the main card file. The regulator power transistor is mounted on the heat-sink assembly.

The dissociator electronics provide for monitoring the oscillator voltage, current, and RF level at the collector terminal. In addition, an optical monitor is incorporated to provide a quantitative and objective measure of the brightness of the discharge. Two integrated photodetector/amplifiers view the discharge through "atomic" and "molecular" interference filters at 6562 Å (the Balmer γ line of atomic hydrogen) and 6013 Å (near a hydrogen-band spectrum), respectively, to provide outputs that can be used to verify proper operation of the dissociator.

Phaselock Receiver

The VLG-11 phaselock receiver is a totally new design, fully utilizing recent improvements in L-band and RF amplifiers. An RF amplifier, with 20-db gain at 1420 MHz, is incorporated in the isolator box. The preamplifier is temperature controlled to within $\pm 0.05^\circ\text{C}$ to help present a more constant input impedance to the isolator output port.

Within the receiver proper (Figure 12), the second preamplifier stage is preceded by an input filter, a wide-band device whose only requirement is to reject the image-frequency noise input at approximately 740 MHz. A double-balanced harmonic mixer follows the second RF amplifier. The local-oscillator frequency for this first mixer is 360 MHz; the third harmonic of 360, 1080 MHz, is generated in the mixing process and beats with the 1420.4-MHz RF from the maser to generate the 240.4-MHz first intermediate frequency. Harmonic mixing is inherently phase stable, although there is a penalty of approximately 12 db in conversion loss. The receiver noise figure is determined by the preamplifiers preceding the first mixer, and the overall noise figure is unaffected by the additional loss. The output of the first mixer is filtered and down-converted again in the second mixer. The second local-oscillator frequency is again 360 MHz, giving a second IF of 19.595 MHz. The 19.595-MHz frequency is amplified and routed to the IF amplifier module, where a third down conversion, to 405 kHz, is accomplished. This IF amplifier has a gain of approximately 60 db and hard-limits the signal in a symmetrical nonsaturating manner.

The output of the amplifier limiter drives the balanced phase detector in the integrator module. The reference for the phase detector is the nominal 405.751685-MHz output of the digital synthesizer. The DC output of the phase detector is amplified and used to tune electronically the 60-MHz voltage-controlled crystal oscillator (VCXO). Auxiliary circuitry in the integrator module indicates the maser signal level by means of a second quadrature phase detector and sweeps the VCXO in the absence of a quadrature output to facilitate initial acquisition.

The 60-MHz output of the VCXO drives an emitter-coupled-logic (ECL) limiter and pulse shaper, which in turn drives a $\times 3$ multiplier. The output of the $\times 3$ multiplier, at 180 MHz, is amplified to drive a full-wave frequency doubler, the 360-MHz output of which is used as the local-oscillator signal for both the first and the second mixer. The output of the pulse shaper also serves as the input to ECL digital dividers, which provide 20- and 5-MHz outputs. Both the $\times 3$ multiplier and the $\div 3$ dividers are referenced to the extremely fast transition of the ECL pulse shaper to improve phase coherence between the local-oscillator signals and the 5-MHz outputs. To improve the phase stability further,

the pulse shaper, the frequency multipliers, and the digital dividers are housed in an oven module, which holds the internal temperature to $\pm 0.1^\circ\text{C}$.

Two independent 5-MHz output channels are separately buffered for isolation; a third channel provides ECL-compatible 5.0 MHz to the digital-synthesizer reference input. Separate buffered outputs are available at 60 and 360 MHz for special applications.

The receiver box is mounted on the front panel of the maser cabinet just below the control assembly. The oven module can be seen at the lower right in Figure 4, behind the bracket that holds the oven-control electronics. The RF module (input plate) is on the upper left, and the shielded modules containing the printed-circuit boards for the IF amplifier, the integrator, the 360-MHz distribution amplifier, and the two 5-MHz buffers are just below the input plate on the lower left.

In keeping with the general philosophy of the VLG-11 design, all monitoring and power functions are provided by the maser system. The receiver requires no operator controls nor adjustments.

Digital Synthesizer

The VLG-11 digital synthesizer is also a totally new design. Previous SAO maser systems employed commercial synthesizers, which were bulky and heavy and required 65 to 100 watts of 110-volt AC power. The VLG-11 synthesizer is quite compact and draws only 3 watts from the 28-volt DC maser power supply. The synthesizer provides a resolution of 0.001 Hz over a tuning range of 405,750.000 to 405,753.999 Hz.

The primary requirements imposed on the synthesizer are low phase noise, low phase drift, and a very low level of spurious outputs at frequencies within 10 kHz of the carrier. The system designed for this purpose is shown in block diagram form in Figure 13. Each of two subloop circuits generates two variable digits, using digital divide-by-N phaselock loops. A simplified block diagram of subloop 1 is shown in Figure 14. A single LSI CMOS integrated circuit provides the $\pm N$ programmable counter and digital phase-detector functions. The reference frequency for the phase detector is relatively high, 10 kHz, so that very good reference-frequency suppression, better than -70 db, is possible at the output of the VCO.

The outputs of the two subloops are combined in the 5.5-MHz loop, which locks a third VCO to the sum of the two subloop frequencies plus 5.0 MHz. The 550-kHz output of the loop is translated to 5.750 kHz in the translator, using frequencies derived from the 5.0-MHz synthesizer reference, and fed to the 4.05-MHz loop. This circuit locks a 4.05750-MHz crystal oscillator to the sum of the output of the translator and the 4.0-MHz signal derived directly from the reference. The output of the crystal oscillator is divided by 10 to yield the 405,750.000- to 405,753.999-Hz drive signal for the receiver phase detector.

The short-term stability of the synthesizer has been measured by offsetting two synthesizers by 1.2 Hz and measuring the period of the beat output from a low-noise phase detector. The results are shown in Table 1.

The synthesizer box is mounted on the front of the maser below the receiver. All circuitry is mounted on shielded printed-circuit modules that plug into a back

Table 1. Short-term stability of VLG-11 digital synthesizers. (The 1.42 GHz is the equivalent noise when the synthesizer is used in conjunction with the receiver.)

Average time (sec)	Allan variance	
	At 405 kHz	At 1.42 GHz
8.3	1.2×10^{-12}	3.4×10^{-16}
83.3	2.2×10^{-13}	6.3×10^{-17}
833.3	5.9×10^{-14}	1.7×10^{-17}

plane. The modules can be operated on extender cards for maintenance and troubleshooting.

Monitoring of the synthesizer lock-loop status is provided on the maser's monitor panel. A special high-resolution counter and LED readout is included on the monitor panel to display the four least-significant digits of the synthesizer output frequency. Thumb-wheel switches for setting these four digits are located on the maser control panel. Operating power for the synthesizer is furnished from the power supply for the main maser.

Monitoring and Control

The VLG-11 incorporates an extensive monitoring capability. Four front-panel mounted meters, each with an eight-position selector switch, are located on the monitor panel to provide quick-look monitoring of 32 functions. These include all main power-supply voltages, all heater voltages, hydrogen dissociator operating conditions, receiver/synthesizer signal levels, and phaselock control voltages.

The monitor panel also includes the small four-digit counter and LED readout for displaying the output frequency of the synthesizer. The counter is independent of the synthesizer and serves as a check on the proper operation of the synthesizer. Six LED lamps on the panel provide a quick-look indication of the status of the main power, thermal, pressure, pump, and dissociator oscillator subsystems.

In addition to the analog metering, the VLG-11 provides an internal telemetry system. Thirty-one channels of voltage or current data are normalized to a standard 0- to +5.0-volt full scale, buffered, and brought out to a 61-pin cylindrical test connector on the front panel. This analog output is designed to simplify the interfacing of a digital data-processing system to the maser for continuous recording of critical operating parameters. Five monitoring thermistors are also accessible through this connector to permit convenient measurements of the temperature of each vacuum tank and oven thermal zone.

Performance Data

Only a relatively limited amount of performance data for the VLG-11 is currently available.

Short-term stability. The short-term stability data taken to date are shown in Figure 15, with comparable VLG-10A data shown for reference. VLG-11 data for averaging intervals greater than 100 sec are not yet available.

Environmental stability. A comprehensive environmental test program for the VLG-11 masers will begin at SAO in July 1977; thermal, magnetic-field, and barometric-pressure sensitivity will be evaluated.

Preliminary test results indicate that the torri-spherical shield design increases the magnetic-shielding factor by approximately 1.5. To verify this improvement, a large Helmholtz coil is under construction, designed to permit testing under carefully controlled conditions. Similarly, a special chamber is being built at SAO to provide barometric-pressure testing capabilities over an ambient-pressure range of +0.0 to +2.0 inches Hg.

Acknowledgments

This work was supported in part by Contract N00014-76-C-0569 from the Office of Naval Research and Contract 375-6026/E from the Max-Planck Institute.

References

1. Kleppner, D., Goldenberg, H., and Ramsey, N., Theory of the hydrogen maser. *Phys. Rev.*, vol. 26, pp. 603-615, 1962.
2. Vessot, R. F. C., Levine, M. W., Mueller, L., and Baker, M., Design of an atomic hydrogen maser system for satellite experiments. In *Proceedings of the 21st Annual Symposium on Frequency Control*, U.S. Army Electronics Command, Ft. Monmouth, N.J., pp. 512-542, 1967.
3. Vessot, R. F. C., and Levine, M. W., Performance data of space and ground hydrogen masers and ionospheric studies for high accuracy frequency comparison between space and ground clocks. In *Proceedings of the 28th Annual Symposium on Frequency Control*, U.S. Army Electronics Command, Ft. Monmouth, N.J., pp. 408-414, 1974.

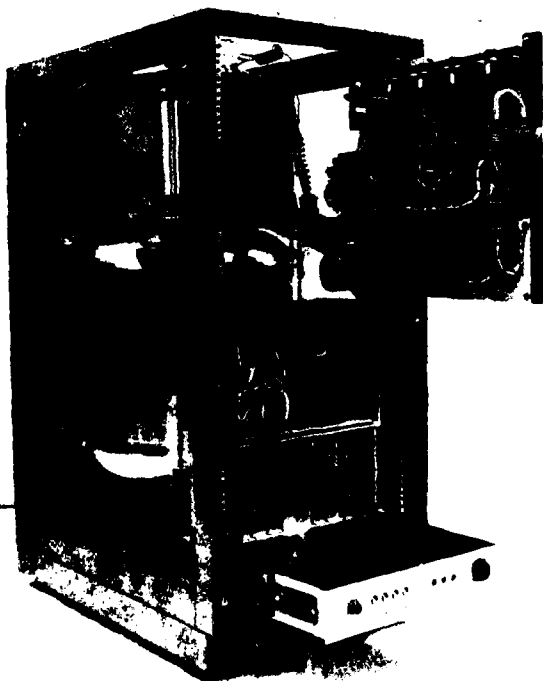


FIGURE 2. VLG-11 HYDROGEN MASER WITH PANELS OPEN FOR SERVICING.

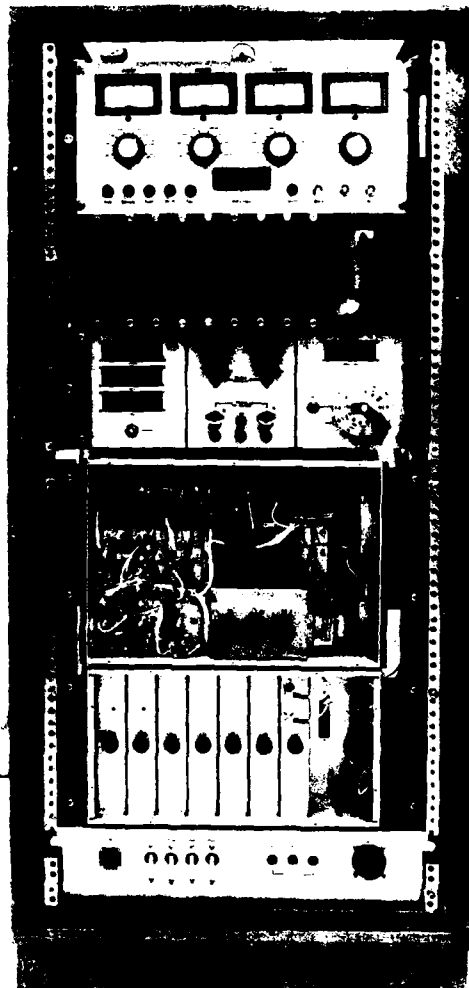


FIGURE 1. FRONT VIEW OF THE VLG-11 HYDROGEN-MASER FREQUENCY STANDARD.

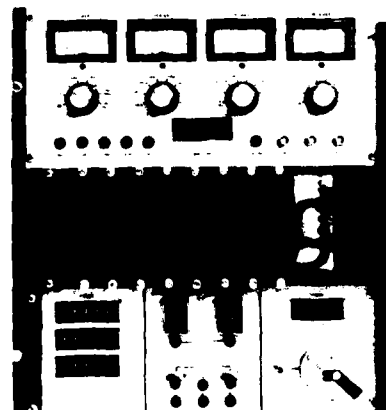


FIGURE 3. CONTROL ASSEMBLY SHOWING THE SHIELDED MODULES IN THE CARD FILE.

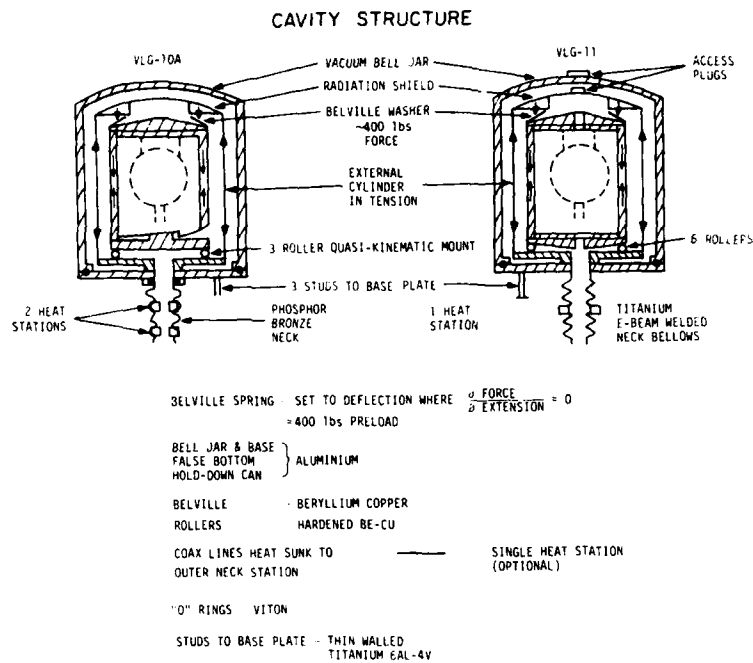
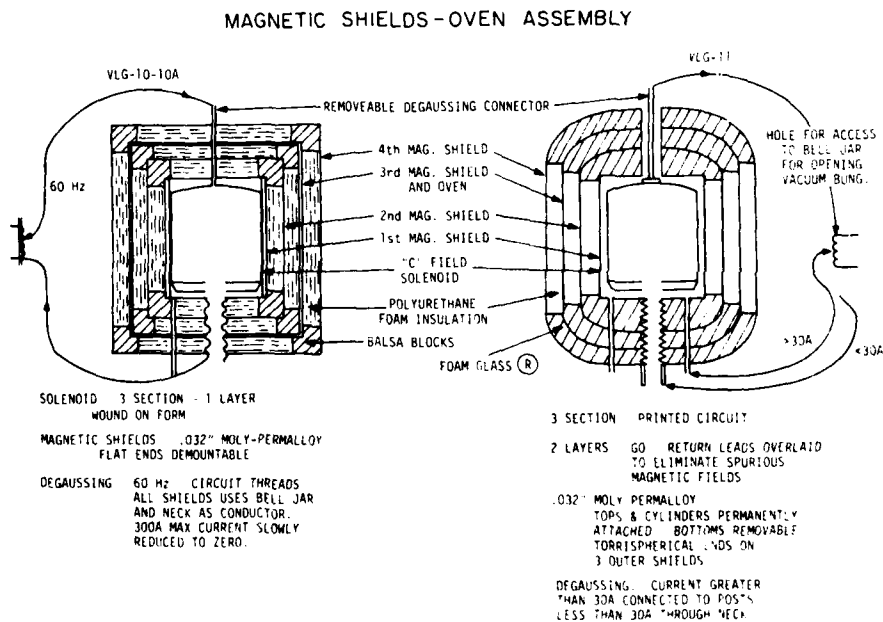


FIGURE 7. CAVITY STRUCTURE AND COMPARISON WITH VLG-10.



(R) PITTSBURGH-CORNING

FIGURE 8. MAGNETIC-SHIELD-OVEN ASSEMBLY AND
COMPARISON WITH VLG-10.

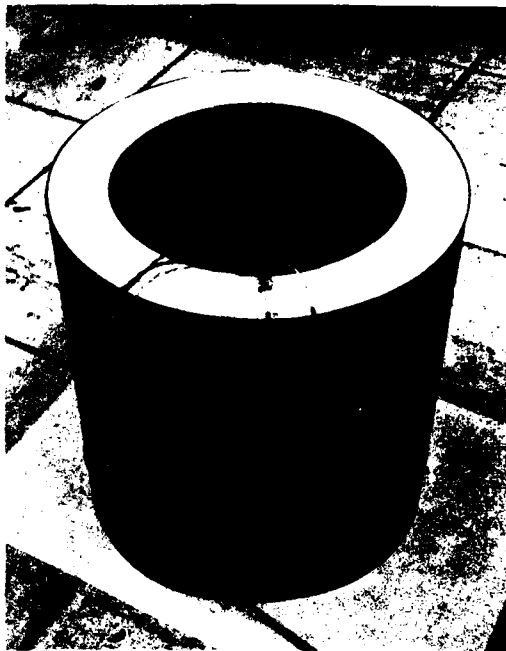


FIGURE 9. PRINTED -CIRCUIT SOLENOID ASSEMBLY.

THERMAL CONTROL

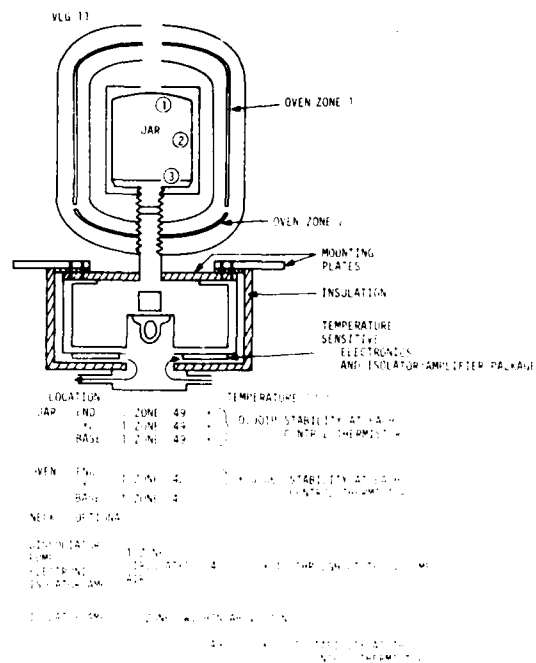
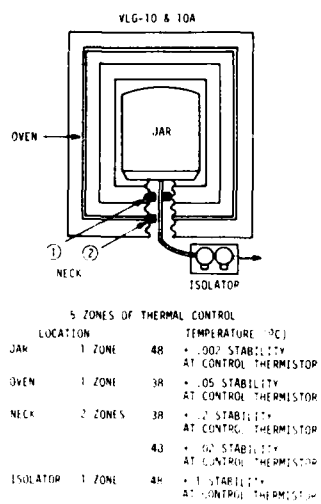


FIGURE 10. THERMAL CONTROL SYSTEM AND COMPARISON WITH VLG-10

VLG-10-10A AND VLG-11 ION PUMP MANIFOLD AND DISSOCIATOR-STATE SELECTOR

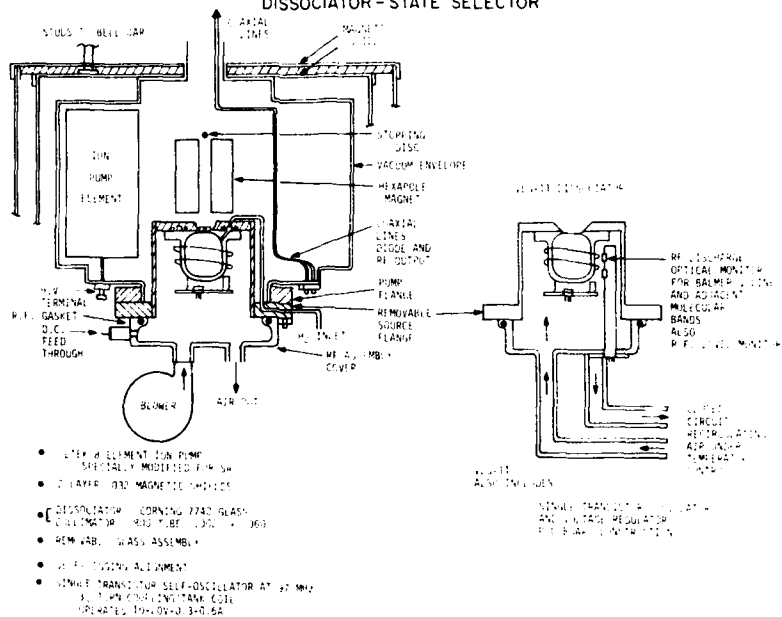
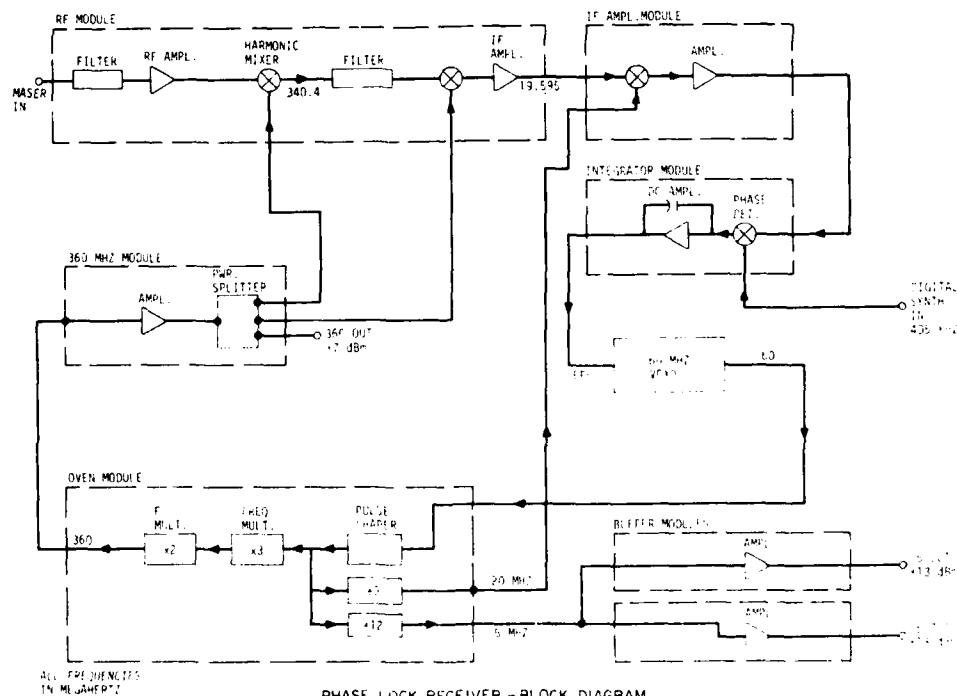


FIGURE 11. ION-PUMP SOURCE ASSEMBLY AND COMPARISON WITH VLG-10.



PHASE LOCK RECEIVER - BLOCK DIAGRAM

FIGURE 12. BLOCK DIAGRAM OF THE PHASELOCK RECEIVER.

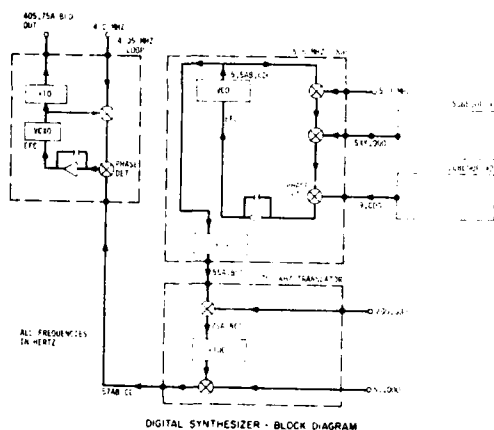


FIGURE 13. BLOCK DIAGRAM OF THE DIGITAL SYNTHESIZER.

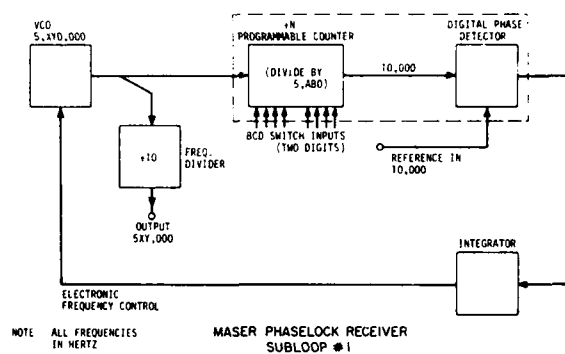


FIGURE 14. BLOCK DIAGRAM OF SUBLOOP 1 IN THE DIGITAL SYNTHESIZER.

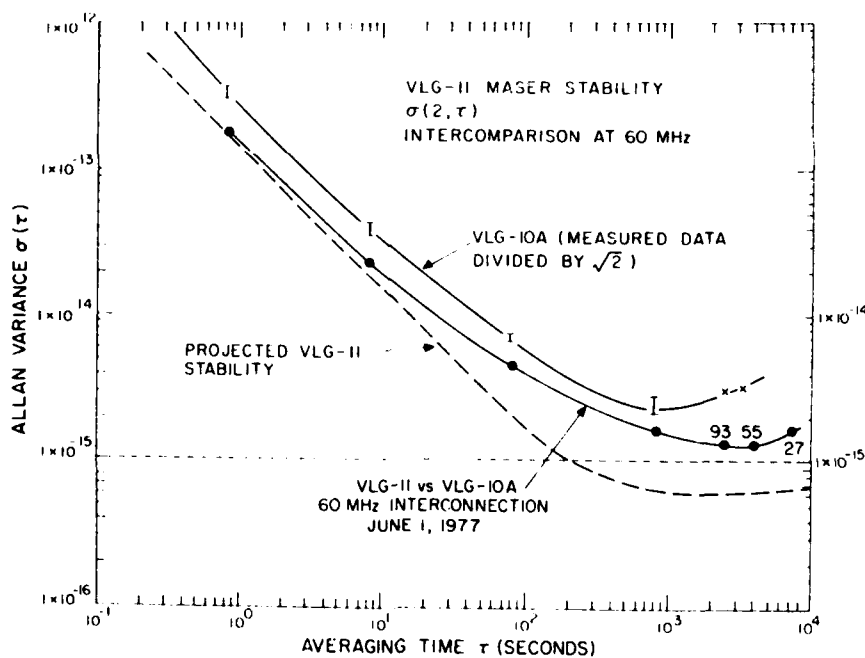


FIGURE 15. PRELIMINARY STABILITY DATA ON THE VLG-11 MASER.

ZEEMAN EFFECTS ON H AND Rb MASERS

Giovanni Busca, Jean-Yves Savard, Silvana Rovea, Jacques Vanier
Université Laval, Québec, Canada

Michel Desaintfuscien, Pierre Petit, Claude Audoin
Université de Paris, Orsay, France

Summary

A theory describing the H maser frequency and amplitude response to coherent Zeeman perturbations is presented. Measurements performed on a H maser for different magnetic fields are found to be in good agreement with the theory. For uniformly populated Zeeman sublevels and low magnetic fields this theory predicts that the Zeeman perturbation add a pure broadening term to the ordinary H maser frequency equation. This prediction is confirmed by preliminary experimental results. Cavity tuning by Zeeman effect on uniformly populated Zeeman sublevels is also investigated experimentally. Finally results relative to the Zeeman effects on the Rb maser are given.

Introduction

The Zeeman effect on H and Rb masers has been studied experimentally and theoretically by several authors (1,2,3,4,5). However, only partial theories of the phenomenon has been given so far, despite the practical importance of these studies which allow to predict or control the effect of magnetic field inhomogeneities and magnetic field fluctuations on the long term frequency stability of the maser. A more complete theoretical and experimental study of Zeeman effects on H maser is presented here and is analysed from a practical point of view.

Zeeman Effects on H maser - Theory

This work is an extension of a similar work done by Andresen (1). However we take into account the full spin-exchange phenomenon in presence of Zeeman coherence (6) and we make the calculation for unrestricted maser amplitude. Details of the calculations are given in Ref. (7) Here the most important results are summarized. The hydrogen atomic levels of interest are presented in Figure 1. The Zeeman sublevels are identified respectively by numbers 1,2,3; the maser transition takes place between levels 2 and 4.

Maser Physical Parameters Calculation

The Zeeman line-width (w_{12} , w_{23}) and the maser line-width (B_{42}) are expressed by:

$$\begin{aligned} w_{12} &= r + \gamma_2 + \frac{1}{4}u(1-p) \\ w_{23} &= r + \gamma_2 + \frac{1}{4}u(1+p) \\ B_{42} &= r + \gamma_2 + \frac{1}{4}u \end{aligned} \quad (1)$$

where r is the bulb escape rate, γ_2 is the density independent line-width, taking into account relaxation by wall and magnetic inhomogeneities; $u/2 = n\tau$ represents the longitudinal spin-exchange relaxation rate and $p = \rho_{11} - \rho_{33}$ is the atomic polarization.

The population difference for the oscillating maser are calculated from fundamental parameters and

in the absence of Zeeman perturbations:

$$\begin{aligned} D_{12}^0 &= a_{12} + \left(\frac{2}{(B + \frac{u}{4})B_{42}} - \frac{u/2}{(B + \frac{u}{4})B_{42}^2} \right) (D_{24}^0)^2 (X_{42}^0)^2 \\ D_{23}^0 &= a_{23} + \left(-\frac{2}{(B + \frac{u}{4})B_{42}} + \frac{u/2}{(B + \frac{u}{4})B_{42}^2} \right) (D_{24}^0)^2 (X_{42}^0)^2 \\ D_{24}^0 &= \frac{r\rho_{22}^B}{(b + \frac{u}{4}) + 4\frac{(X_{42}^0)^2}{B_{42}}} \end{aligned}$$

where the source terms a_{12} and a_{23} are expressed by:

$$\begin{aligned} a_{12} &= \frac{r.B.(d_{12} + \frac{u}{4B}d_{23})}{(B + \frac{u}{4})(B - \frac{u}{4})} + \frac{\frac{u}{4}r^2d_{13}}{(B + \frac{u}{4})(B - \frac{u}{4})} \\ a_{23} &= \frac{r.B.(d_{23} + \frac{u}{4B}d_{12})}{(B + \frac{u}{4})(B - \frac{u}{4})} - \frac{\frac{u}{4}r^2d_{13}}{(B + \frac{u}{4})(B - \frac{u}{4})} \end{aligned} \quad (3)$$

In this eq. d_{12} and d_{23} are beam population difference:

$$\begin{aligned} d_{12} &= \rho_{11}^B - \rho_{22}^B \\ d_{23} &= \rho_{22}^B - \rho_{33}^B \end{aligned} \quad (4)$$

The oscillating maser transition rate $(X_{42}^0)^2$

$$= \frac{\mu_0^2 H^2}{h^2} \quad \text{in the absence of Zeeman perturbations is:}$$

$$(X_{42}^0)^2 = \frac{\rho_{22}^B 4\pi\mu_0^2 \text{In} Q_c}{4h\nu_c} - \frac{B_{42}}{4(B + \frac{u}{4})} \quad (5)$$

where μ_0 is the Bohr magneton, I is the total atomic flux in the beam, η is the filling factor, Q_c the loaded cavity quality factor and v_c the cavity volume. Equation 5 is equivalent to the more familiar expression of the maser power as a function of the oscillating parameter q and atomic flux (8).

The main results of this calculation is that the population difference between two given Zeeman sublevels depends on the relative flux into these levels and also, due to the spin-exchange relaxation, on relative flux into the other level. Knowledge of the previous oscillating maser parameters allows the calculation of the maser amplitude and frequency response to Zeeman perturbations. The Zeeman intensity (X_{42}^0) is a scale factor and can be adjusted for fitting the experimental results.

Maser Frequency

We introduce first the notation. Zeeman coherences are shifted by spin-exchange interaction, by the

amount:

$$\begin{aligned}\Omega'_{12} &= \Omega_{12} - \frac{1}{4}vD_{23}^0 \\ \Omega'_{23} &= \Omega_{23} - \frac{1}{4}vD_{12}^0\end{aligned}\quad (6)$$

Consequently the detuning between the external coherent Zeeman excitation frequency ω and the shifted Zeeman transition frequencies, in the limit of nearly tuned cavity, are expressed by:

$$\begin{aligned}\omega - \Omega'_{12} &= \delta'_{12} = \omega - \Omega_{12} + \frac{1}{4}vD_{23}^0 \\ \omega - \Omega'_{23} &= \delta'_{23} = \omega - \Omega_{23} + \frac{1}{4}vD_{12}^0\end{aligned}\quad (7)$$

Introducing the lorentzian shape, associated with Zeeman absorption:

$$L_{ij} = \frac{\omega_{ij}}{\omega_{ij}^2 + (\delta'_{ij})^2} \quad (8)$$

and the associated dispersion:

$$F_{ij} = \frac{\delta'_{ij}}{\omega_{ij}^2 + (\delta'_{ij})^2} \quad (9)$$

it is convenient to define a generalised absorption type curve given by:

$$A_{ij} = \frac{\omega_{ij} + X_{42}^2 L_{ij}}{(\omega_{ij} + X_{42}^2 L_{ij})^2 + (\delta'_{ij} + X_{42}^2 F_{ij})^2} \quad (10)$$

a generalized dispersive type curve given by:

$$F_{ij} = \frac{\delta'_{ij} + X_{42}^2 F_{ij}}{(\omega_{ij} + X_{42}^2 L_{ij})^2 + (\delta'_{ij} + X_{42}^2 F_{ij})^2} \quad (11)$$

and a combined function of both A_{ij} and F_{ij} :

$$C_{ij} = A_{ij}L_{ij} + F_{ij}F_{ij} \quad (12)$$

For simplicity we report in the following the theoretical results relative to a tuned cavity.

The maser frequency (ω_m), including the Zeeman effects, is given by the expression:

$$\begin{aligned}\omega_m - \Omega'_{24} &= \Delta\omega_{s.e.} + \Delta\omega_t + \Delta\omega_p + \frac{2Q}{\Omega_{24}}\Delta\omega_c(B_{42} + b) \\ &+ \frac{\epsilon_n}{T_h}\end{aligned}\quad (13)$$

where Ω_{24} is the unperturbed atomic H frequency corrected for the second order magnetic field dependence; $\Delta\omega_{s.e.}$ is the spin-exchange frequency shift in presence of Zeeman perturbations, $\Delta\omega_t$ is an asymmetry parameter dependent term, $\Delta\omega_p$ is a polarisation dependent term, b represents the Zeeman line broadening and ϵ_n/T_h is a recently introduced spin-exchange shift term (2).

The spin-exchange shift ($\Delta\omega_{s.e.}$) expression is:

$$\Delta\omega_{s.e.} = -\frac{1}{4}v(D_{24}^0 + D_{24}') \quad (14)$$

where D_{24}^0 is the population difference in absence of

Zeeman perturbations, as given by Eq. 2 and D_{24}' represents the change induced by the Zeeman perturbations expressed by:

$$D_{24}' = \frac{2}{B + \frac{u}{4}} \left(4A_{12}X_{12}^2D_{12}^0 - A_{23}X_{23}^2D_{23}^0 + \frac{C_{12}X_{12}^2}{B_{42}} \right. \\ \left. (X_{24}^0)^2D_{24}^0 + \frac{C_{23}X_{23}^2}{B_{42}}(X_{42}^0)^2D_{24}^0 \right) \quad (15)$$

In the absence of Zeeman perturbations $\Delta\omega_{s.e.}$ is proportional to the maser full line-width (16). This statement is not generally true when Zeeman perturbations are present. In this last case the maser oscillation condition is:

$$\frac{R}{B_{42} + b} = \frac{rhv_c}{4\pi\mu_o^2 I_{tot} nQ} \quad (16)$$

$$\text{where: } R = D_{24}^0 + D_{24}' + C_{12}X_{12}^2D_{12}^0 - C_{23}X_{23}^2D_{23}^0 \quad (17)$$

which corresponds to a spin-exchange shift given by:

$$\begin{aligned}\Delta\omega_{s.e.} &= -c\lambda(B_{42} + b) + \frac{1}{4}nv\lambda(C_{12}X_{12}^2D_{12}^0 \\ &- C_{23}X_{23}^2D_{23}^0)\end{aligned}\quad (18)$$

where the usual notation for v ($v = nv\lambda$) has been introduced and c is a density independent parameter

($c = \frac{v_c v_h}{16\pi Q n v_b \mu_o}$). Eq. (18) shows that only for uni-

formly populated Zeeman sublevels the spin-exchange shift in presence of Zeeman perturbation is proportional to the maser full line-width.

The asymmetry parameter dependent term:

$$\Delta\omega_t = F_{12}X_{12}^2 - F_{23}X_{23}^2 \quad (19)$$

is the sum of two near centered and opposite dispersion type curves. It is independent of the Zeeman sublevels populations. This term represents a direct contribution of the coherent Zeeman perturbation which consists of two opposite pulling effects on the maser frequency. Neglecting the spin-exchange shift these pulling cancel themselves perfectly in the limit of zero magnetic field.

The polarisation dependent term $\Delta\omega_p$ is given by:

$$\begin{aligned}\Delta\omega_p &= -\frac{8\pi\mu_o^2 I nQ}{rhv_c} \left[\frac{L_{12}\delta'_{12}X_{12}^2D_{12}^0}{(\omega_{12} + (X_{42}^0)^2L_{12})^2 + (\delta'_{12} + (X_{42}^0)^2F_{12})^2} \right. \\ &+ \left. \frac{L_{23}\delta'_{23}X_{23}^2D_{23}^0}{(\omega_{23} + (X_{42}^0)^2L_{23})^2 + (\delta'_{23} + (X_{42}^0)^2F_{23})^2} \right] \quad (20)\end{aligned}$$

This is the dominant term for an ordinary maser operating with no uniform Zeeman sublevels populations and become smaller when the polarisation p is decreased. The term vanishes rigorously only for uniform Zeeman sublevels populations i.e. for $D_{12}^0 = D_{23}^0 = 0$. As a conclusion the maser frequency equation useful for low magnetic fields and uniform Zeeman sublevels populations is:

$$\omega_m = \Omega_{24} + \left(\frac{2Q}{\Omega_{24}} \Delta\omega_c - c\lambda \right) (R_{42} + b) + \frac{\epsilon_h}{T_h} \quad (21)$$

This equation illustrates a zero total Zeeman induced frequency shift and pure Zeeman broadened line.

This equation predicts the existence of a cavity setting for which the maser frequency is independent of both Zeeman intensity and Zeeman frequency. This cavity setting is named the Z.I.F.S.

Maser Power

Maser power variations (Δp_m) due to the effect of coherent Zeeman perturbations is expressed by:

$$\begin{aligned} \frac{4\Delta p_m}{B_{42}(B + \frac{u}{4}) + 4(X_{42}^0)^2} &= \frac{1}{D_{24}^0} \left(\frac{2A_{12}}{B + \frac{1}{4}u} + C_{12} \right) X_{12}^2 D_{12}^0 \\ &- \frac{1}{D_{24}^0} \left(\frac{2A_{23}}{B + \frac{1}{4}u} + C_{23} \right) X_{23}^2 D_{23}^0 + \frac{2}{(B + \frac{1}{4}u)B_{42}} \left(C_{12} X_{12}^2 (X_{24}^0)^2 \right. \\ &\left. + C_{23} X_{23}^2 (X_{24}^0)^2 \right) - \frac{b}{B_{42}} \quad (22) \end{aligned}$$

where $\Delta p_m = (X_{42}^0)^2 - (X_{42})^2$ and the broadening b is:

$$b = A_{12} X_{12}^2 + A_{23} X_{23}^2 \quad (23)$$

Eq. (22) shows that the maser power is increased by Zeeman perturbations of frequency Ω_{12} if $D_{12}^0 > 0$. On the contrary the maser power is decreased by Zeeman perturbations of frequency Ω_{23} if $D_{23}^0 > 0$. This is a usual situation for masers having one magnetic state selector. The broadening term b appearing in this Eq. does not depend on the Zeeman sublevels populations. Consequently it contributes, in all situations, to the maser power decrease. The broadening b is the dominant term when uniform population between Zeeman sublevels is obtained.

Experimental Results

H Maser Experimental Results

The experimental results have been obtained on H masers having 6 magnetic shields and provided with a RF coil at the exit of the state selector magnet for mixing Zeeman populations.

Measurements have been performed for both high and low magnetic fields. The agreement between the theory and the experiment is satisfactory and illustrated in Figure 2. Maser frequency changes and relative maser power changes induced by the Zeeman perturbations are plotted against the Zeeman frequency. The points are the experimental values and continuous curves are calculated theoretically using the parameter values shown in figure 2.

We may notice that:

1. The maser power dip, center frequency, does not occur at the exact Zeeman transition frequency due to complicated effects associated with the doublet structure of the Zeeman line. The frequency separation of the two lines is $\approx 0.387\text{Hz}$.

2. The zero crossing of the maser frequency dispersion-like curve occurs for Zeeman frequency lower

than the maser power dip, center frequency. Theoretically this frequency difference is a function of the maser power and cavity frequency.

3. A structure in the maser power apparently uncorrelated with the Zeeman frequencies appears. Fig. 3 shows how the structure in the power curve arises. The effect of each Zeeman line has been calculated separately. Index 1 refers to the 1-2 Zeeman transition and index 2 refers to the 2-3 Zeeman transition.

In Fig. 4 a comparison for Zeeman frequencies around 26.5 KHz is shown. This measurement illustrates the fact that Zeeman perturbation can contribute either to an increase or a decrease in the maser power depending on both flux distribution and the Zeeman frequency.

In Fig. 5 a typical comparison at low magnetic field is shown. Both theoretical and experimental values of the frequency displacement between the power dip, center frequency, and the dispersion-like curve, center frequency, become negligible.

Mixing the Zeeman Sublevels Populations

Uniform Zeeman populations can be obtained experimentally using a R.F. coil placed at the exit of the magnetic state selector. In our measurements the coil exciting frequency was approximately 907 KHz, the beam diameter was about 4 mm and the interaction region length was approximately 2 cm. Fig. 6 shows how the maser dispersion-like frequency curve is changed into a straight line by proper Zeeman mixing. Similar results have been obtained previously by V. Reinhardt (3). The dispersion around the straight line is 5×10^{-13} . In scale B the maser frequency shifts have been expanded by a factor of 10. The measurement averaging time was ≈ 12 sec. The quoted dispersion is approximately 1 order of magnitude greater than the maser stability for the same averaging time. Noise appearing on the straight line seems to be due to instabilities associated with the mixing technique. Fig. 7 shows the maser frequency for a cavity detuning of ≈ 600 Hz. The effect of Zeeman broadening appears evident. The minimum cavity detuning detectable in the experimental conditions illustrated, was approximately 150 Hz.

Rb Maser Results

87 Analytical calculations are not available for the Rb maser due to the complexity of the problem which requires a solution of 8×8 density matrix equations. However the results obtained on H maser can be transferred qualitatively to the Rb maser.

Fig. 8 shows typical experimental results obtained for the Rb maser. Also for the Rb maser, the zero crossing of the maser frequency dispersion-like curve occurs at a Zeeman frequency lower than the maser power dip, center frequency.

Curves 1 and 3 corresponds to almost opposite cavity detuning. In the curve 1 zero maser frequency shift occurs at the exact Zeeman frequency corresponding to the minimum power. This is the usual situation when a resonant coherent Zeeman tuning is used.

Conclusion

Some practical suggestions of the present work are the use of the Zeeman effect on the maser power for magnetic state selection efficiency measurement and the possibility of Z.I.F.S. for the tuning of the H maser. We have calculated here the maser response to a monochromatic perturbation. The maser response to any Zeeman

man perturbation can be obtained by the convolution of the present response function with the given Zeeman perturbation spectral density. In particular the effect of DC magnetic field inhomogeneities can be derived.

Finally from a more fundamental point of view, improvement of the Zeeman mixing technique will allow measurements of the spin-exchange parameter ϵ_h , the output frequency difference between a s.e. tuned maser and a Z.I.F.S. tuned maser being ϵ_h times the maser line width.

Bibliography

- (1) H.G. Andresen, Z. Physik 210, 113, (1968).
- (2) S.B. Crampton and H.T.M. Wang, Phys. Rev. A 12, 1305, (1975).
- (3) V.S. Reinhardt and H.E. Peters, An improved method for measuring the magnetic inhomogeneity shift in H masers, Proceeding, 29th Annual Symposium on Frequency Control, U.S. Army Electronic Command, Fort Monmouth, N.J. pp 357-361, (1975).
- (4) J. Vanier and R.F.C. Vessot, IEEE J. Q.E. 2, 391, (1966).
- (5) C. Cohen Tannoudji and J. Dupont-Roc, Phys. Rev. A 5, 968, (1972).
- (6) P.L. Bender, Phys. Rev. 132, 2154, (1963).
- (7) J.Y. Savard et al., Zeeman effects on H maser, to be published.
- (8) D. Kleppner et al., 138, A972, (1965).
- (9) M. Desaintfuscien et al., Journal de Phys. 36, 281, (1975).
- (10) S.B. Crampton, Phys. Rev. 158, (1967).

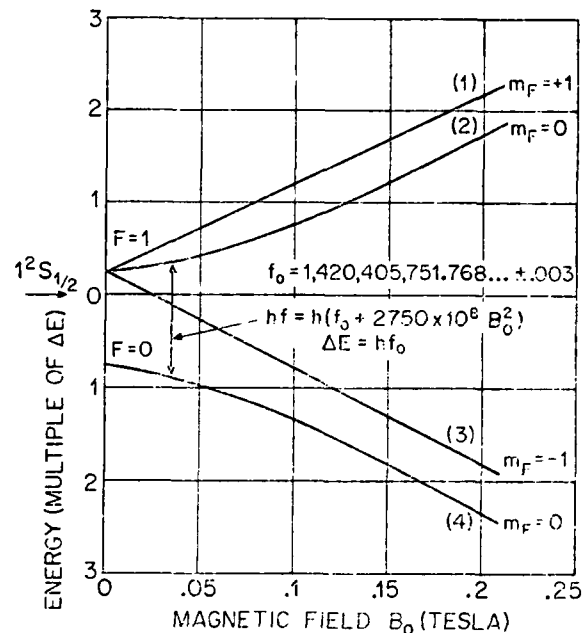


Fig. 1: Energy levels of H atoms;

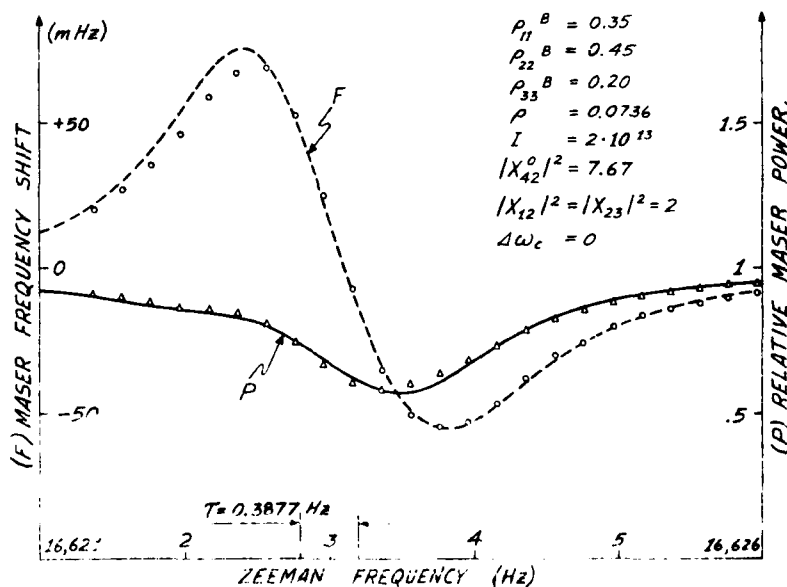


Fig. 2: typical Zeeman effects on H maser for high magnetic fields;

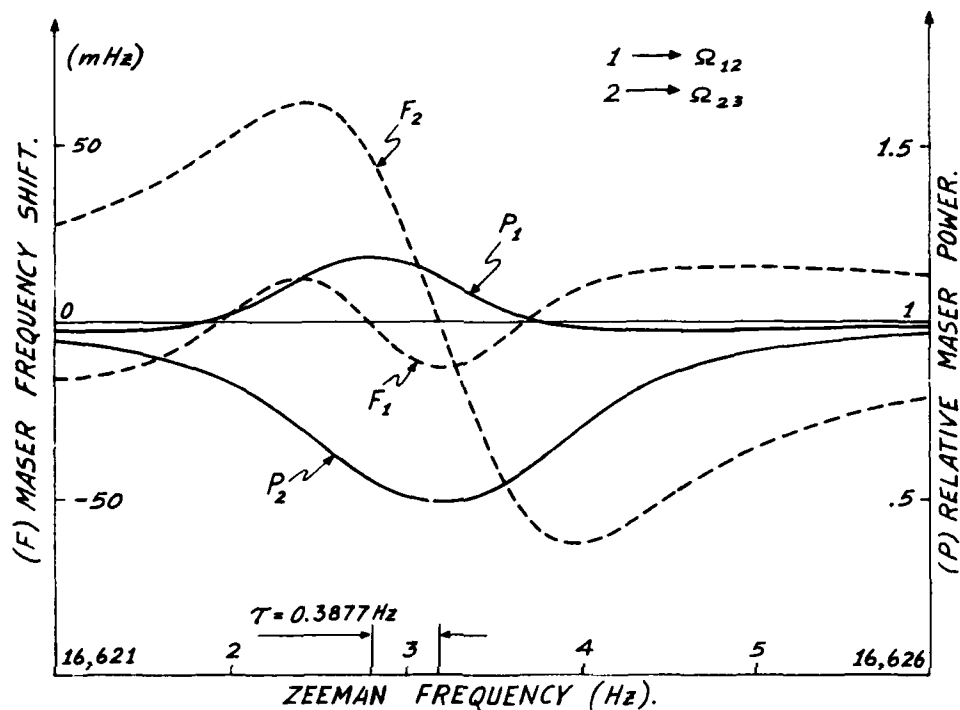


Fig. 3: theoretical explanation of the results given in Fig. 2;

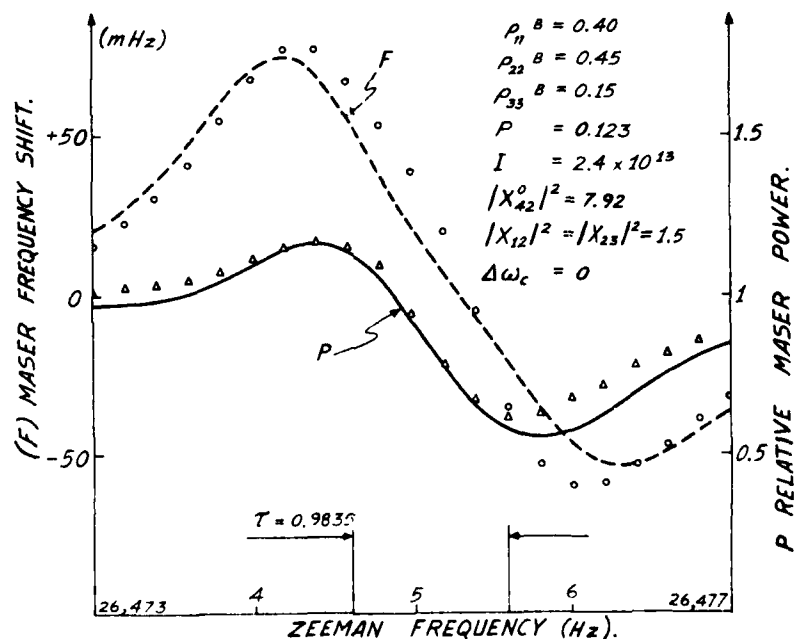
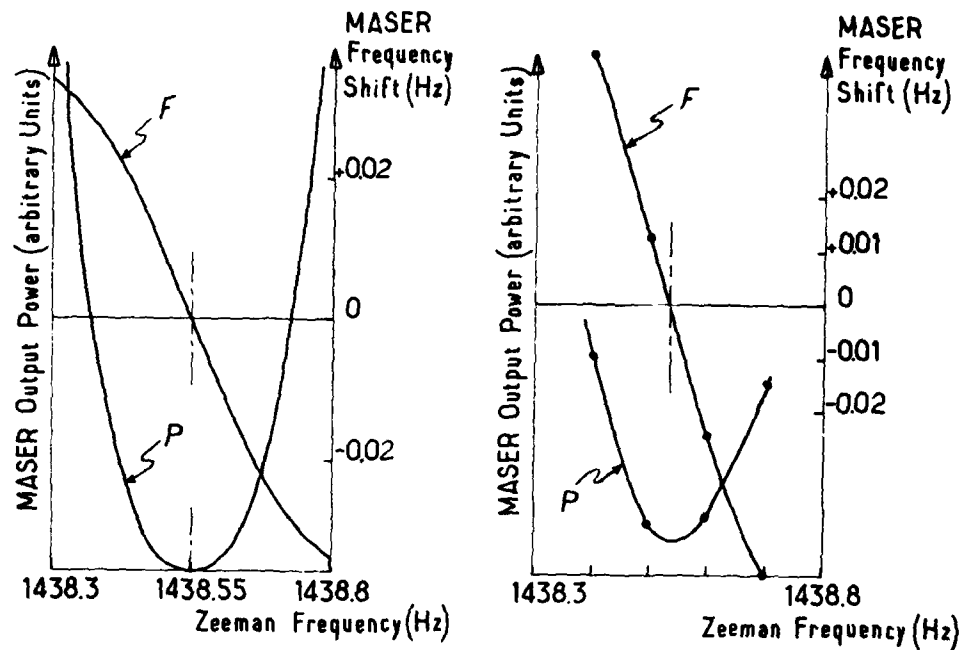


Fig. 4: typical Zeeman effects on H maser for high magnetic fields;



Theoretical Curves

Experimental Results

Fig. 5: typical Zeeman effects on H maser for low magnetic field;

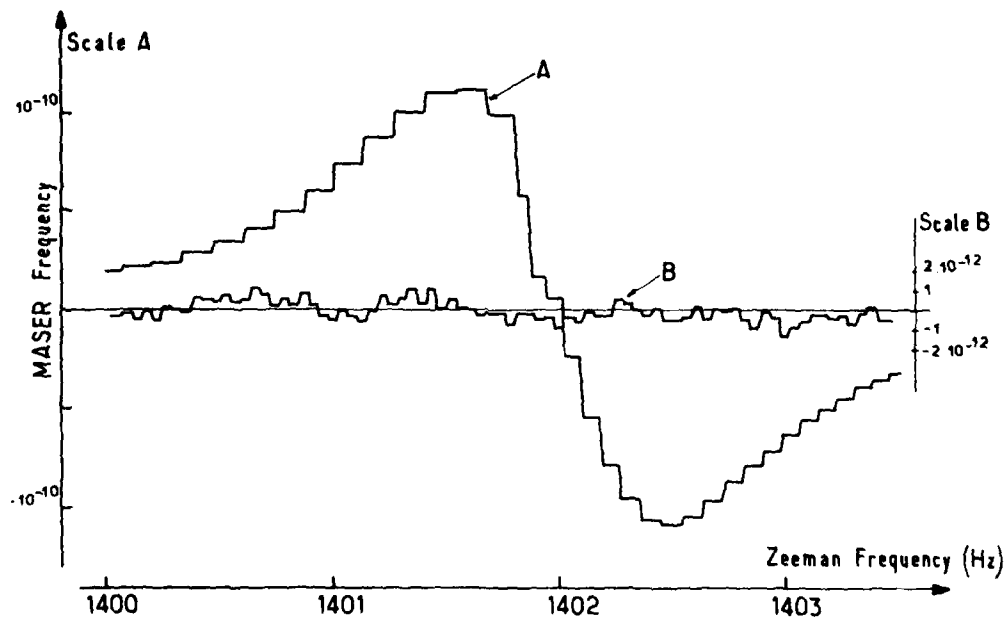


Fig. 6: maser frequency vs Zeeman frequency; curve A: normal maser operation; curve B: uniform Zeeman populations;

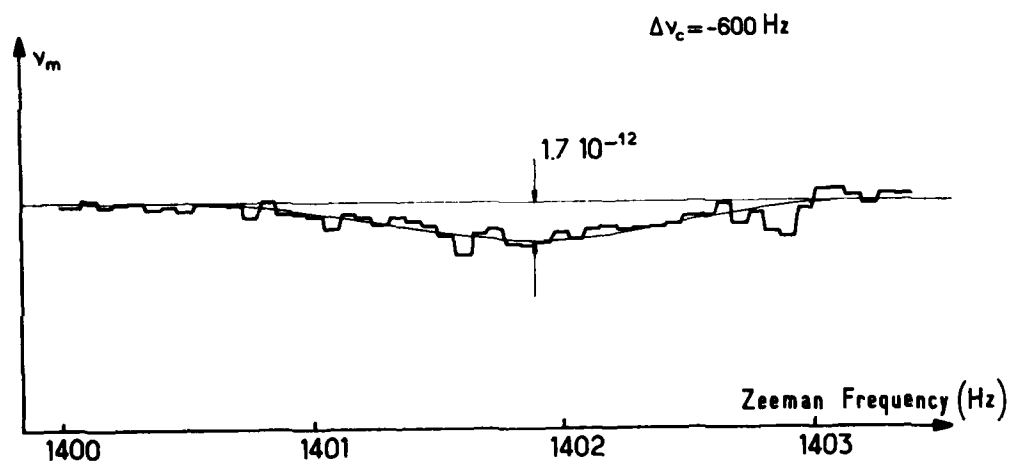


Fig. 7: maser frequency vs Zeeman frequency for detuned cavity;

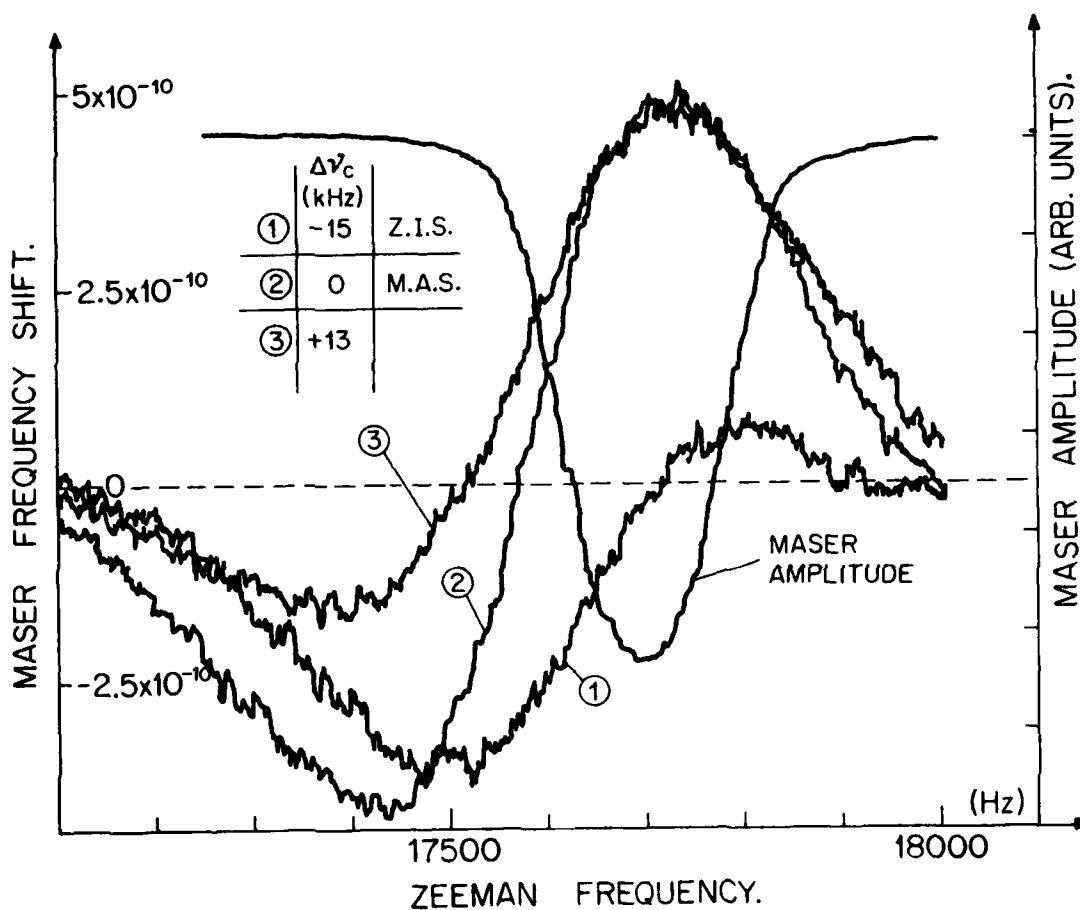


Fig. 8: typical Zeeman effects in Rb maser.

DEVELOPMENT OF A CESIUM BEAM CLOCK FOR SATELLITE APPLICATION

James George and Alfred I. Vulcan
Frequency Electronics, Inc.
New Hyde Park, New York

Summary

A new Cesium Beam Clock has been developed whose design concept focuses on those parameters necessary for a space environment. The characteristics which were optimized for space use also resulted in a device which has application to airborne and ground based usage with stringent environmental conditions. These characteristics are low weight, low power consumption, radiation hardened, rugged construction, and long life. This paper discusses the general design concept and presents test data taken on two prototype units.

Introduction

Prototype resonators are constructed of 10", 12" and 14" length. In addition 20" tubes are presently in production for laboratory use. The 14" length was selected as the best compromise between frequency stability and tube life. The projected lifetime for the 14" tube is seven years minimum, with the projection based upon data gathered during 1 1/2 years of laboratory operation.

Some of the features used in the new resonator are as follows:

1. Indirectly heated Niobium target with a 10 section electron multiplier is utilized. The signal-to-noise ratio is 55 to 60 db as compared to 40 db for spectrometer detectors.
2. Improved shielding allows a small size.
3. An improved magnetic structure operating at 0.8 alpha for velocity selection gives a higher signal level as compared to standard tubes operating at 0.3 alpha.
4. Improved Cesium oven for zero gravitational environment using a metallic sponge structure preventing Cesium migration from the collimator.
5. Ruggedized construction permits operation to vibration levels of 28 Grms.

Design

Figure 1 is a table showing the basic environmental characteristics of the new Cesium Beam Resonator. Broad temperature range operation is obtainable through the use of a compensated waveguide structure and thermistor compensation networks for the multiplier diodes. The compensation permits usable operation over the temperature range of -55°C to +71°C with a coefficient of $1.3 \times 10^{-13}/^{\circ}\text{C}$. The temperature

variation is due mainly to thermal expansion of the C-Field shield.

Figure 2 shows the Ramsey curve for the 14" Cesium Beam Resonator. The symmetry of the first order field independent state with the main peak centered indicates a linear magnetic field and good magnetic shield integrity. The high peak-to-valley ratio is an indication of good phase coherence in the microwave tee.

Figure 3 is a block diagram of the entire Cesium Beam Standard. The power dissipation of the tube including the oven, primary loop, multiplier, and detector is 8.5 watts. At -60°C an additional 3.5 watts is required for the Cesium oven. The primary loop has a time constant of 1 second and the secondary loop time constant is 20 seconds. An ovenized 14.59 MHz oscillator feeds a x630 multiplier chain consisting of a x18 and x35 Step-Recovery-Diode multiplier. The x35 output at 9.19 GHz feeds the resonator input.

A low noise FET amplifier at the tube output converts the detector signal to a usable level.

The Cesium oven controller is a high efficiency device using a pulse width modulation to achieve 82% efficiency. The power supply configuration consists of D.C. converters, switching regulators, and a series regulator in each module. This approach results in extremely low R.F. modulation by line related frequencies. The synthesizer accepts the 14.59 MHz signal and converts it to 5 MHz. The 5 MHz also drives a Time-of-Day clock which presents both a visual indication and a serial TOD output.

Figure 4 is a block diagram for the primary loop. A novel feature of this circuit is an auto lock provision such that in the event of a loss of lock in the satellite environment due to power outage, high radiation inputs, or EMC transients the system automatically reacquires the proper Ramsey peak within 30 seconds. Referring to figure 4 it is seen that the detected output from the FET amplifier drives a synchronous detector and a 2 millisecond delay compensates for the phase shift through the Cesium Beam Resonator. The offset null loop compensates for any long term drift of the integrator circuit so it has negligible effect on frequency.

A loss of lock initiates the sweep generator which slews the VCXO over a range of ± 1 PPM. At the same time the synchronous detector is inhibited and the output feeds a peak detector and hold circuit. Since the side lobes are 70% of the main lobe, the comparator and scaler senses when the main lobe is acquired, enabling the

synchronous detector and locking the system. In actual use the VCXO sweeps through a minimum of 3 peaks and on the return sweep lock is obtained.

Figure 5 is a block diagram of the synthesizer and secondary loop. This circuit accepts the 14.59 MHz output from the primary loop and divides and mixes internally generated frequencies to synthesize 5 MHz. A programmable divider with a division ratio of 1,450 to 2,050 allows the output frequency to be changed over a range of 1×10^{-8} with a resolution of 1.6×10^{-13} . The 5 MHz oscillator in the secondary loop has extremely low phase noise and excellent short term stability and in the free-running condition the output signal has an aging rate of 2×10^{-11} per day. The secondary loop attack time is 20 seconds.

Various alarm conditions indicating loss of lock or other frequency anomalies are outputted to indicate improper system operation.

Specifications

The specifications for the satellite frequency standard are as follows:

<u>PARAMETER</u>	<u>VALUE</u>
Warm-Up	14 Minutes
Accuracy	$\pm 1 \times 10^{-11}$
Reproducibility	3×10^{-12}
Stability	$\pm 1.6 \times 10^{-13}$
Long Term Stability	$\pm 1 \times 10^{-11}/3$ years
Short Term Stability	$1.2 \times 10^{-9}/100$ microseconds $1 \times 10^{-12}/\text{Sec.}$
Jitter	25 psec. rms
Pulse Advance	100 nsec steps 1 sec maximum
Auto Sync	± 0.5 microseconds
Battery Capacity	2 Hours at 25°C 1 Hour at $+55^\circ\text{C}$ and -30°C
Signal to Noise Ratio	60 db
Line Width	1100 Hz
Beam Tube Life	4 Years

This data is compiled for two prototype frequency standards for an 18 month period.

In order to determine if any life limiting characteristics are inherent in the design of the standard, signal-to-noise data and frequency have been measured on a weekly basis. This data is presented in figures 6 and 7. The signal-to-noise ratio

which measured approximately 60 db during the first month of operation for both units has fallen off gradually to a level of approximately 55 db where it is remaining constant. This indicates that the gettering capacity and the structural integrity of the resonators is proper.

Figure 7 shows long term frequency stability for both resonators referred to LORAN C. A total excursion of 4×10^{-12} peak-to-peak is seen around a fixed offset of 2×10^{-13} . This data is well within the original design limits.

Physical Characteristics

The Cesium Beam Standard as packaged in a 19" rack panel assembly is shown in figure 8. Modular construction is used with each assembly being RFI filtered and independently replaceable. A tabulation for the various subassemblies is shown in the figure. Figure 9 is the front panel view of the laboratory version of the standard. The panel height is 5.25" and the weight including batteries is 60 lbs. When packaged for D.C. operation the overall weight is 41 lbs. An extremely ruggedized construction is used since the standard is designed to operate under severe military requirements of shock and vibration. The assembly is completely RFI gasketed.

Conclusion

This paper describes recent developments in the manufacture of a new Cesium Beam Frequency Standard that has excellent electrical and mechanical characteristics and is applicable to both space and ground based use. The design is radiation hardened and constructed to meet severe environmental requirements.

CESIUM BEAM FREQUENCY STANDARD

MODEL FE-5440A

ENVIRONMENTAL SPECIFICATIONS

PARAMETER	VALUE
Vibration	$0.5 \text{ g}^2/\text{Hz}$ random
Temperature	- 30° C to + 65° C
Radiation Effect	$1 \times 10^{-12}/10^4$ Rads
Magnetic Field	0 to 25 oersteds
External Magnetic Field	$\pm 2 \times 10^{-12}/2$ Gauss
Reliability	10,000 Hours MTBF
Weight	57 lbs. A.C. Operation 41 lbs. D.C. Operation Only

FIGURE 1. ENVIRONMENTAL SPECIFICATIONS

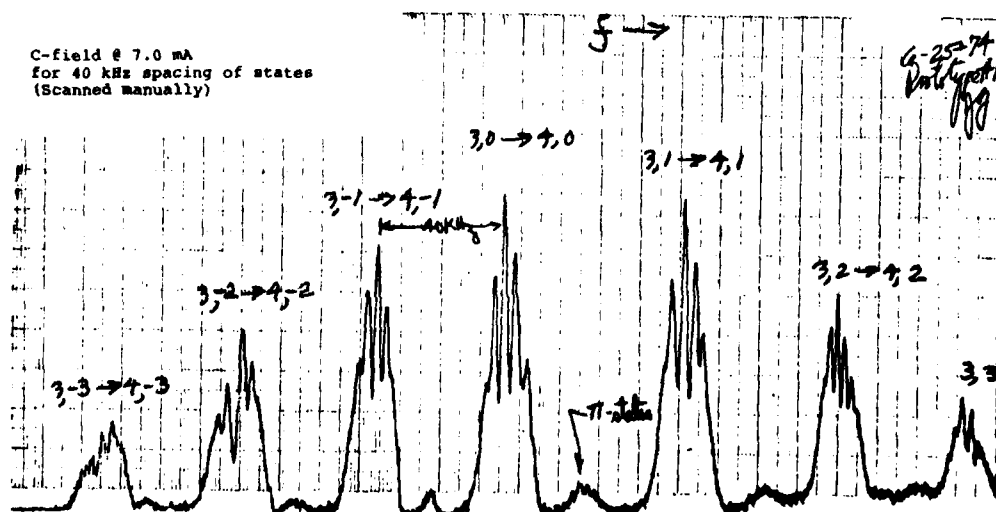


FIGURE 2. RAMSEY CURVE

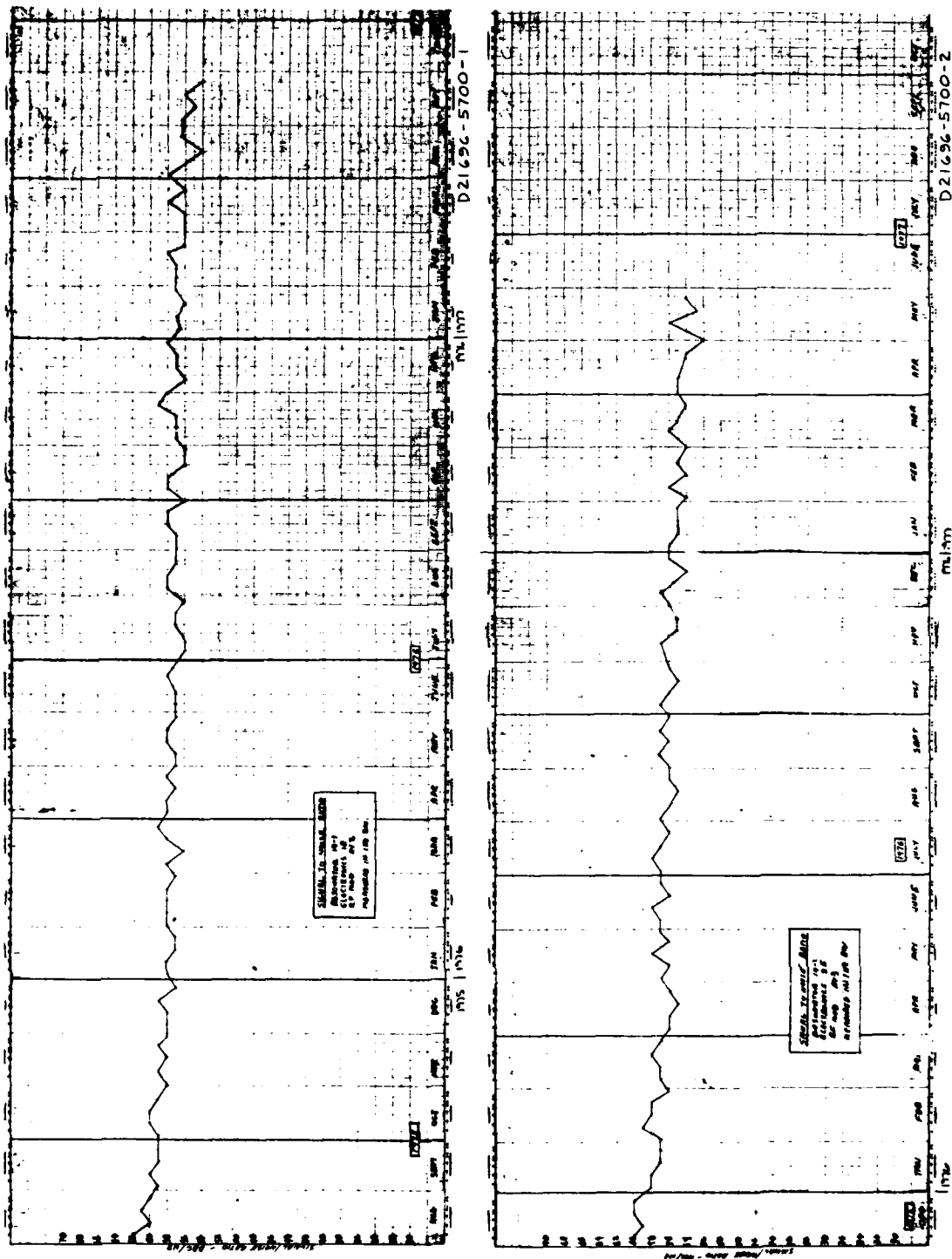


FIGURE 6. SIGNAL-TO-NOISE RATIO

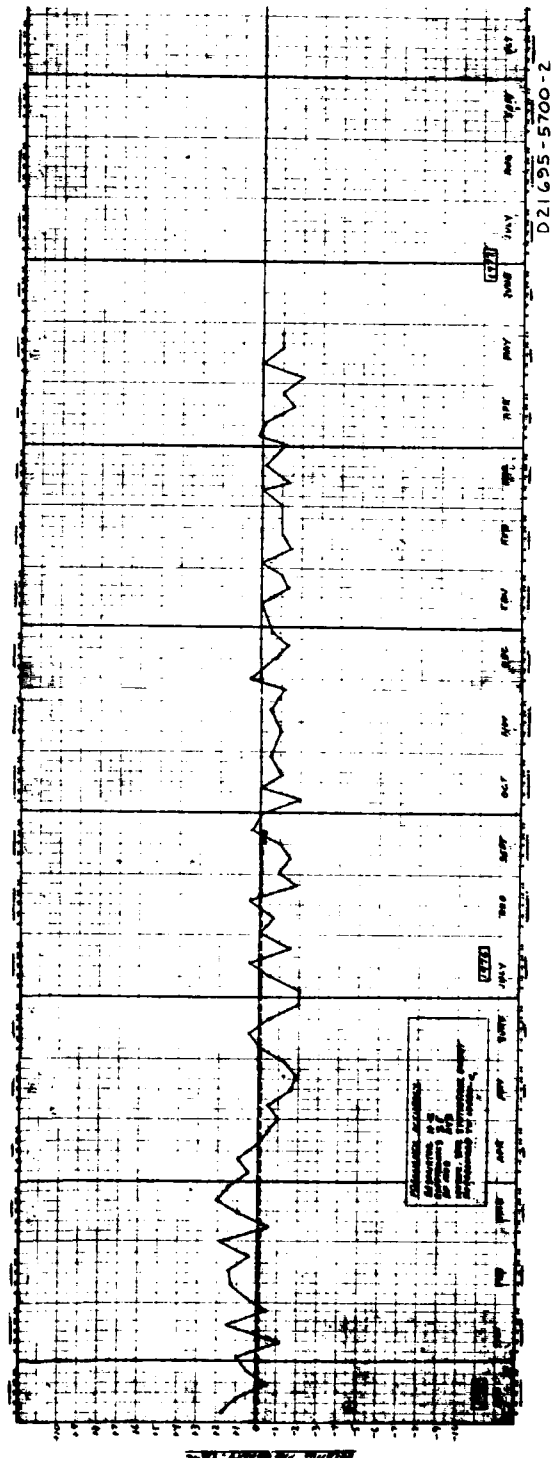
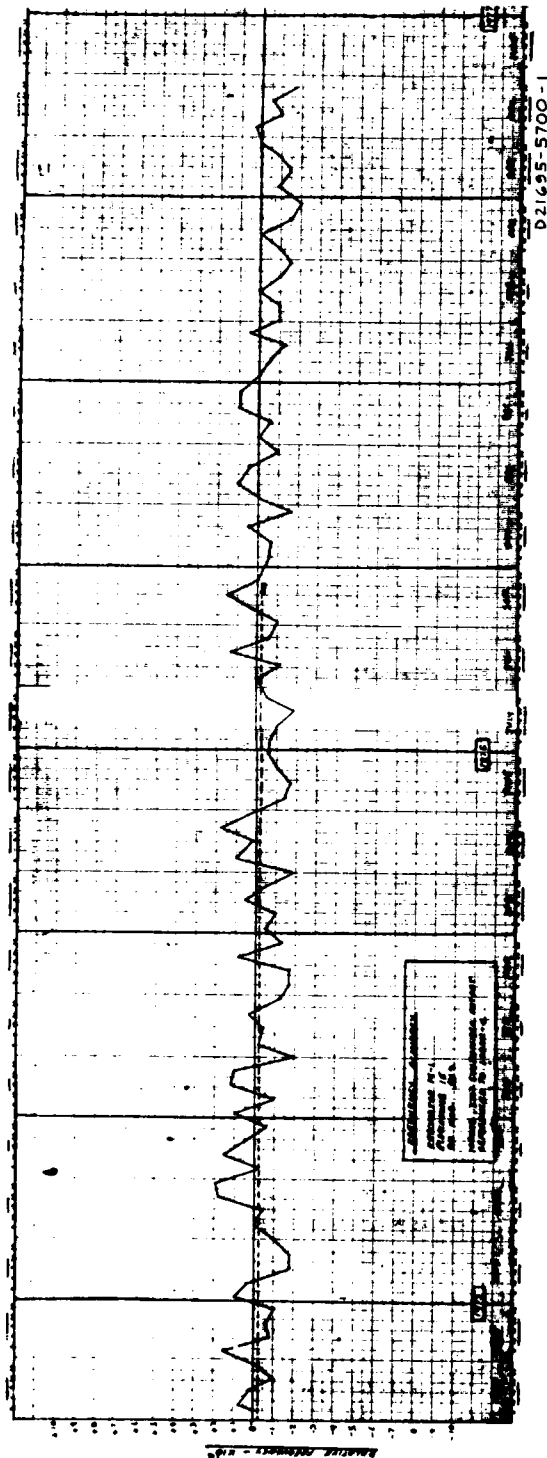
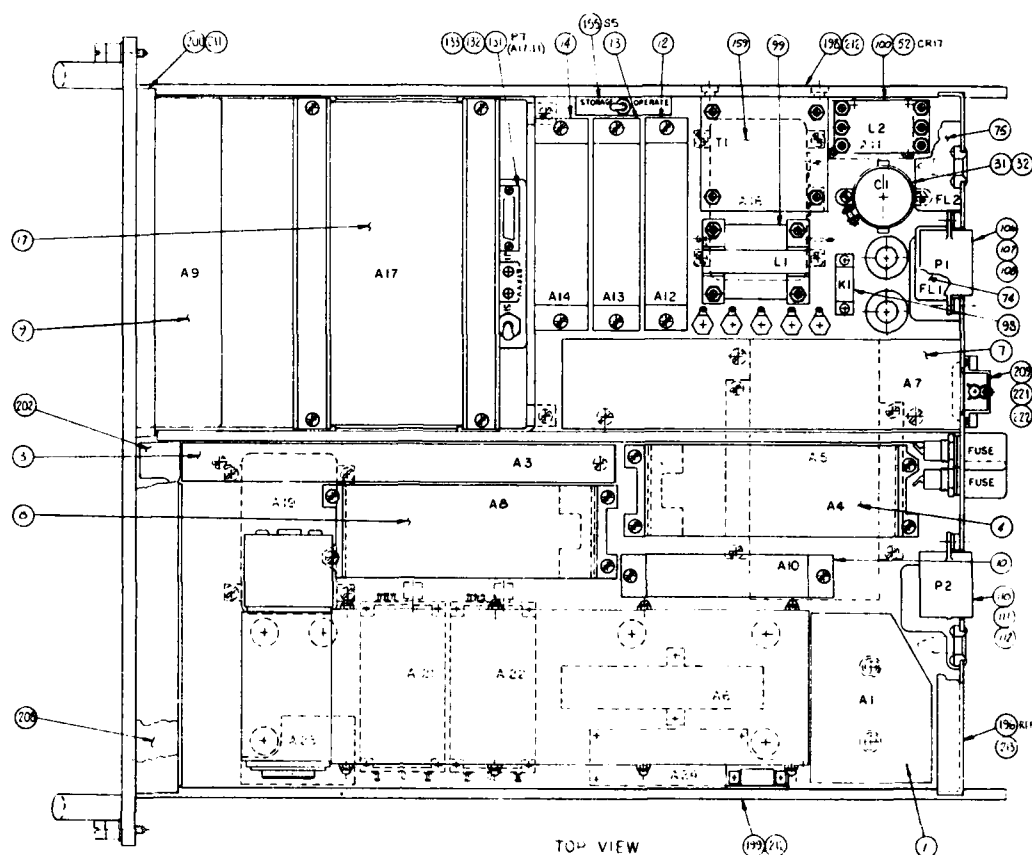


FIGURE 7. STABILITY



LIST OF MODULES

DESCRIPTION

1. PORTABLE PROGRAM CLOCK, PE5440A:

A0 PORTABLE PROGRAM CLOCK, PE-5440A

2. CESIUM BEAM RESONATOR:

A1 CESIUM BEAM RESONATOR ASSEMBLY

A1A1 CESIUM BEAM TUBE

A1A2 PREAMPLIFIER PC ASSEMBLY

A1A3 "C" FIELD SUPPLY PC ASSEMBLY

A1A4 CESIUM OVEN CONTROL PC ASSEMBLY

A2 NOT USED

3. PRIMARY LOOP:

A3 PRIMARY LOOP MODULE

A4 OCVCXO MODULE, PRIMARY, 14.59⁺ MHz

A5 MULTIPLIER MODULE, TIMES 630

A6 WAVEGUIDE ASSEMBLY

4. SECONDARY LOOP:

A7 SYNTHESIZER MODULE

A8 OCVCXO MODULE, SECONDARY, 5 MHz

A9 REAL TIME-OF-DAY CLOCK MODULE

A10 GENERATOR MODULE, 1 MHz, 3 MHz

5. SYSTEM POWER SUPPLIES:

A11 POWER SUPPLY PC ASSEMBLY

A12 SWITCHING REGULATOR MODULE, + 18 Vdc

A13 DC/DC CONVERTER MODULE:

A14 BATTERY CHARGER AND CROSSOVER MODULE

A15 NOT USED

A16 SWITCHING REGULATOR MODULE, + 5 Vdc

A17 BATTERY POWER SUPPLY MODULE, + 28 Vdc

6. CESIUM BEAM SUPPLIES:

A18 NOT USED

A19 SWITCHING REGULATOR MODULE, + 6 Vdc
(HOT WIRE SUPPLY)

A20 NOT USED

A21 POWER SUPPLY, HIGH VOLTAGE, NEGATIVE,
ELECTRON MULTIPLIER

A22 POWER SUPPLY, HIGH VOLTAGE, POSITIVE,
VAC-ION

A23 REGULATOR PC ASSEMBLY, ELECTRON MULTIPLIER

A24 VAC-ION REGULATOR AND CESIUM BEAM
INTERLOCK PC ASSEMBLY

7. METERING:

A25 AMPLIFIERS PC ASSEMBLY METER DRIVER

A26 LOGIC PC ASSEMBLY, ALARMS

FIGURE 8. ASSEMBLY, CESIUM BEAM CLOCK
INTERNAL VIEW

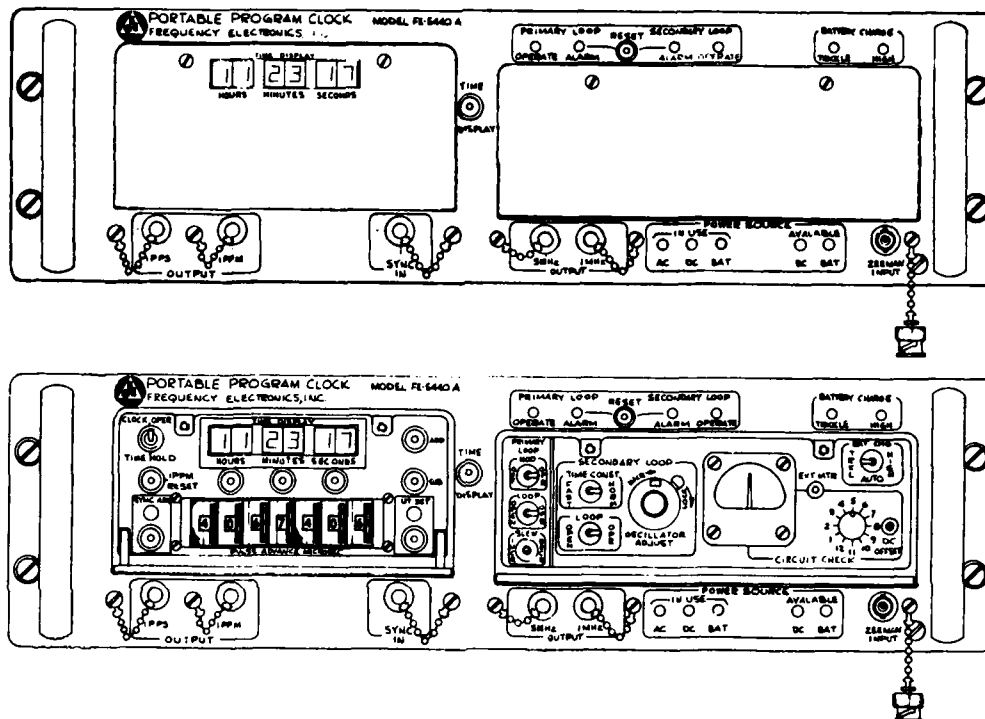


FIGURE 9. FRONT VIEW, CESIUM BEAM CLOCK

NEW CESIUM BEAM FREQUENCY STANDARDS FOR FLIGHT AND GROUND APPLICATIONS

T. K. Gregory
Frequency and Time Systems, Inc.
Danvers, Massachusetts 01923

Summary

The evaluation of the design of a cesium frequency standard for GPS satellite application and the performance of prototype units is described. Cesium frequency standards for GPS satellites must be designed:

1. For lightweight, low power consumption and small size,
2. to withstand shock and vibration from launch,
3. to operate for five years without adjustments in the space environment.

The FTS 1 cesium beam tube has been modified to meet the shock and vibration requirements, and the success of the modification has been demonstrated.

The electrical design approach is conventional but space application requires new or more careful designs of the circuits and sub-systems. Five year life without adjustments place severe requirements on the design of the servo system. The extraction of the small fundamental frequency component from the beam tube output is usually realized with multiple feed-back operational amplifier "active filter" circuits. The stability of such circuits with temperature and time depends on the stability of associated components. The stability of the component values is not likely to be adequate for five years unattended operation in space. Hence, for the space application a commutating filter approach was employed. Unlike active filters, commutative filters do not depend on component values for their characteristics but on a digital signal derived from the quartz oscillator.

To minimize the effects of gain degradation in the integrator during five years in space, a hybrid/analog digital integrator is employed. In addition, the digital portion of the circuit offers a convenient natural interface for the direct digital control of the quartz oscillator when the cesium loop is inactivated; and it is relatively easy to change the effective time constant of the overall integration circuit by remote switching.

Introduction

The design of a cesium beam frequency standard for satellite use was started three years ago under a contract with the Naval Research Laboratories. The first units delivered were prototype units designed to operate in the MTS 2 Spacecraft. At present, two cesium units are installed in the satellite ready for launch. Meanwhile, a second generation of standards designed to meet the requirements of the Navstar-GPS Program have been fabricated. Both the prototype and EDM units, unlike laboratory cesium standards, must survive a severe launch environment; operate unattended for many years; consume minimum amount of the available

spacecraft power, size, and weight; and tolerate the radiation environment. Once on orbit, the environment is quite benign, and performance specifications for the unit are similar to those of a commercial standard.

System Design

The block diagram shown in Figure 1 is generally similar to approaches previously used in commercial units. The system differs mainly in the approach used to implement certain blocks.

The 10.23 MHz output is derived from a 5.115 MHz crystal oscillator which is servo controlled by comparison of a multiplied and synthesized signal from the oscillator to the cesium hyperfine transition. The regulation is accomplished by modulating the 9 GHz signal with an audio signal (f_m) and phase detecting any signal at f_m in the beam tube output.

Servo System

The servo amplifies and filters the modulated output of the beam tube. The filtered signal is phase detected, and the D.C. output integrated to generate the correction signal. In the present system, the usual analog filters and phase detectors have been replaced by digital sampling techniques. The amplified beam tube output is sampled at twice the modulation frequency. The resulting square wave output is phase detected at the modulation frequency by a baseline restorer (Figure 2). The main advantage of this digital approach is improved servo stability. The whole servo system is clocked and does not depend on analog time constants for its accuracy. In addition, no tuning of the system is required to achieve optimum performance. The D.C. error signal generated is then integrated to provide the oscillator control signal. The integration is performed by the hybrid circuit shown in Figure 3. The hybrid integrator consists of an analog integrator which, through upper and lower limit comparators, drives an up/down counter. The output of the counter feeds a R-2R ladder network through relay contacts. Besides acting as low offset switches, the relays also provide immunity to transient by providing magnetic storage of the oscillator control voltage.

There are several unique advantages for this type of circuit. The time constant of the total integrator is the time constant of the analog integrator times the total count of the digital register. Since the current system consists of a 10 bit digital counter and R-2R network, the integrator time constant is:

$$\tau_{HI} = \tau_{AI} \times 1024 = 1024 \cdot R \cdot C$$

For a particular total integrator time constant, the RC time constant can be 1/1024 of the value for a conventional analog integrator. Since the D.C. gain

of the servo system is limited by leakage resistances in the integrator, the ability to keep R small maintains high gain.

The D.C. gain of the GPS system must be considered in two error regions. For errors so small that neither limit will trip, the gain is controlled by the leakage resistance of the integrator. For errors greater than this value, the limits will trip, and the gain is infinite until the error is decreased to below the limit threshold level. If we assume an integrator gain of 10^4 and a system gain of $1 \times 10^{-12} \Delta f/f/MV$, then an error of $1 \times 10^{-12} \Delta f/f$ will produce 10 volts at the output of the analog integrator. For errors greater than $1 \times 10^{-12} \Delta f/f$, the gain is infinite; below this error, the gain is 10^{-12} and the final system offset will be a maximum of $1 \times 10^{-13} \Delta f/f$.

Another advantage of our system is the ability to load the counting register with telemetry information to control the oscillator frequency when the cesium loop is not in use.

In the standard designed for GPS, the integrator has a 30 second time constant and the analog integrator .03 second time constant. The gain of the entire loop is approximately .1, so that, the loop is about 3 seconds. The choice of the loop time constant was controlled by the radiation environment. Under normal circumstances, the satellite environment is very benign, indicating that time constants of hundreds of seconds might be used. However, under transients, quick recovery to lock is required, thus, a short time constant was chosen.

C-Field Supply

The C-Field current supply consists of a relay register similar to that used in the integrator, a stable voltage reference, a 10 bit 2-R2 resistor network and an output amplifier. The stability of the C-Field current is critical to the long term stability of the cesium standard. Errors can be caused by reference drift, resistor network changes and offsets in the amplifier. Tests on several units over the temperature range of -10°C to $+50^\circ\text{C}$ have shown current changes which result in frequency errors of $3 \times 10^{-15}/^\circ\text{C}$. The C-Field can be adjusted through telemetry commands to $4 \times 10^{-13} \Delta f/f$. As in the integrator, relays are used to act as a memory, should a transient occur.

Power Supply

The power supply system shown in Figure 4 consists of a doubly regulated preregulator, 2 inverters, and several output regulators. An L-C filter on the input has a 200 hertz cutoff and attenuates high frequency input and limits the supply ripple reflected back on the power bus.

The first stage of the preregulator boosts the input voltage (22V-30V) to 32V using a switching boost regulator. The second stage is a series pass regulator which produces a well regulated 31 VDC. Measurements on the line regulation of the preregulator show a minimum ripple attenuation of 100 db for frequencies from 30 Hz to 5 KHz.

The regulated 31 volts is fed to two inverters which produce a series of output voltages used for

powering both the beam tube as well as the electronics. Three of the output voltages are regulated. The temperature of the cesium oven is controlled by a switching regulator using a thermistor as feedback. The +15 VDC is generated by a series pass regulator. The 2000 VDC electron multiplier supply uses an A.C. regulator to attenuate modulation related frequencies.

The entire power supply system runs on a clock whose frequency is adjusted to keep switching ripple away from the Zeeman frequency.

System Performance

The performance of the system must be discussed in two categories; (1) survivability of the unit to the launch environment, and (2) operational performance of the unit in the on-orbit environment.

Survivability

Difficulty in achieving survivability is most notably caused by the expected random vibration during vehicle launch. Figure 5 shows qualification and acceptance levels of random vibration for GPS satellite units. The sub-assemblies most affected by these vibration levels are the beam tube and the oscillator. Both of these devices have passed qualification and have been discussed previously²⁻³. The major effort of the mechanical implementation of the standard has been the design of a small lightweight chassis of mounting the tube and oscillator without exhibiting destructive chassis resonances. The chassis shown in Figure 6 has been shaken to qualification level and its resonances examined. While many resonances exist, they all exhibit Q's of three or less at the beam tube and fifteen or less at the oscillator. Tubes and oscillators have been successfully tested in the chassis.

The electronics for the standard do not present a major vibration problem as the techniques of component and board mounting for vibratory environments is well known. The majority of the electronics are contained on four printed circuit boards. Three of the boards; the servo board, the 14.36 MHz synthesizer board, and the dac board (C-field, hybrid integrator, modulation generation) are contained on one side of the standard. The power supply is on a single board mounted on the other side of the standard.

Operational Performance

As was pointed out earlier, the satellite operating environment for the Cs standards is quite benign. The temperature should be held within a few degrees centigrade. There will be no appreciable vibration, and variations in magnetic field will remain within hundreds of milligauss. Under these circumstances, the standard should perform as well as it does in a laboratory environment. The phase noise specification for the GPS Program is shown in Figure 7 along with a typical phase spectrum from the satellite units. The phase noise close to the carrier is higher than that of the oscillator because the loop time constant is short, thus allowing some beam tube degradation of the oscillator spectrum.

The stability of the unit is shown in Figure 8. The very long term stability of the unit has not been measured yet, but based on previous prototype data, it should be below 2×10^{-13} for times greater than 10^5

seconds. The power consumed by the unit is under 25 watts, and the weight under 25 lbs.

Acknowledgement

The authors wish to acknowledge the help of James L. Burkhardt of Frequency & Time Systems, Inc. and James McDade of General Electric Company in this work. We gratefully acknowledge the support of the Office of Naval Research under Contract #N00014-74-C-0061 and of the Defense Nuclear Agency.

References

1. J. White, F. Danzy, S. Falvey, A. Frank, J. Marshall, "NTS-2 Cesium Beam Frequency Standard for GPS", Proc. 8th Annual P.T.T.I. Applications and Planning Meeting, December 1976.
2. E. Graf, L. F. Johnson, R. H. Kern, "A New Compact Cesium Beam Frequency Standard", Proceedings of the 29th Annual Symposium on Frequency Control, 1975.
3. D. A. Emmons, "A New Rugged Low Noise High Precision Oscillator", Proc. 8th Annual P.T.T.I. Applications and Planning Meeting, December 1976.

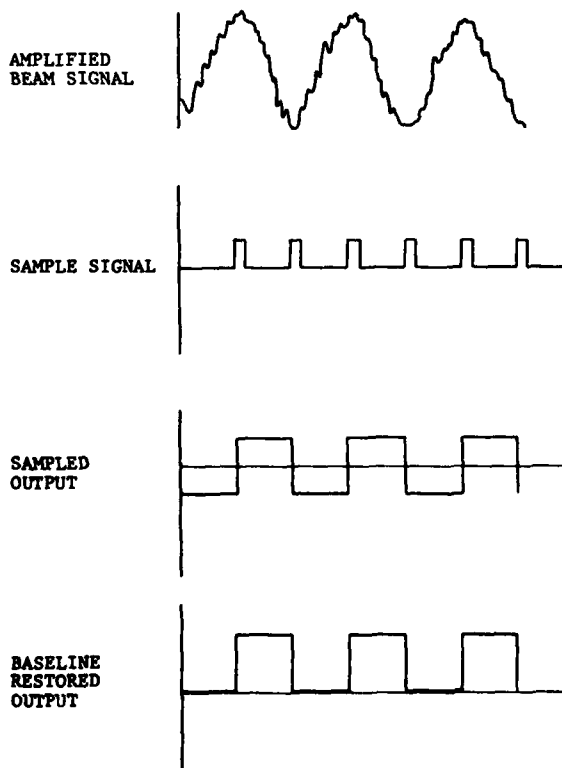


FIGURE 2: SERVO WAVEFORMS

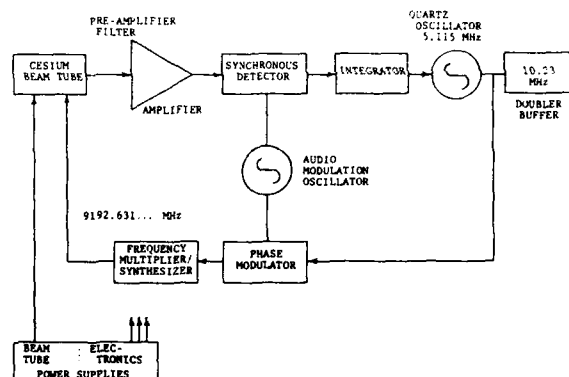


FIGURE 1: CESIUM STANDARD BLOCK DIAGRAM

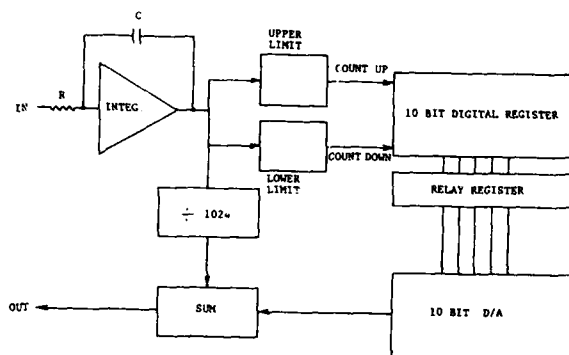


FIGURE 3: HYBRID INTEGRATOR

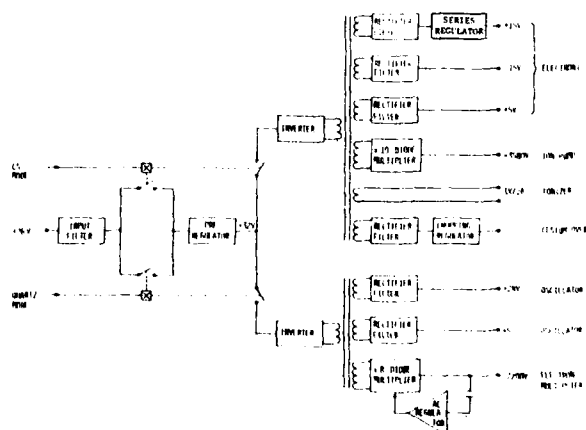


FIGURE 4: POWER SUPPLY

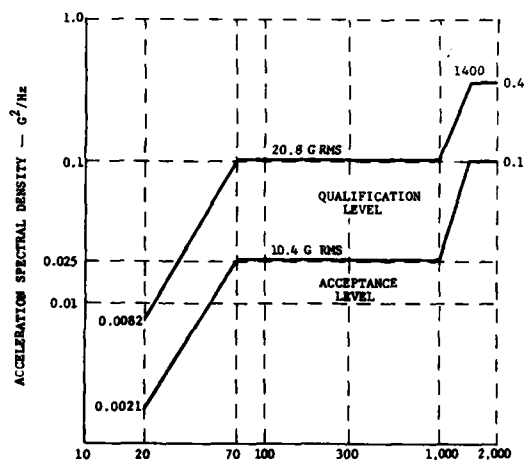


FIGURE 5: VIBRATION LEVELS

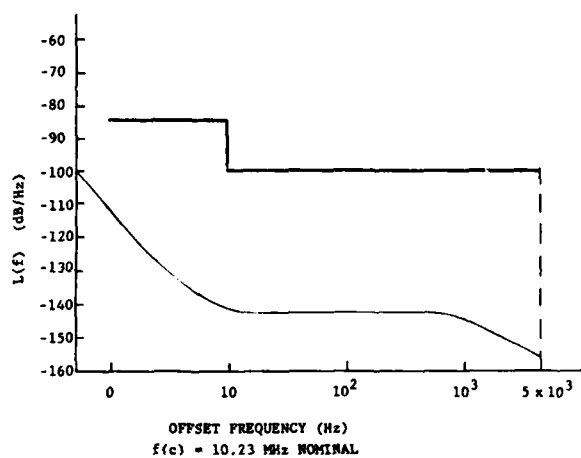


FIGURE 7: PHASE NOISE VERSUS FREQUENCY OFFSET

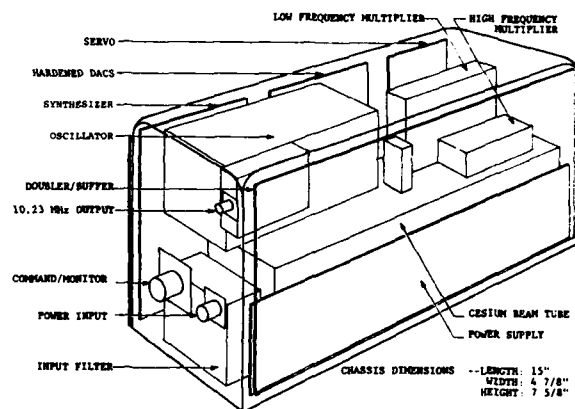


FIGURE 6: CHASSIS

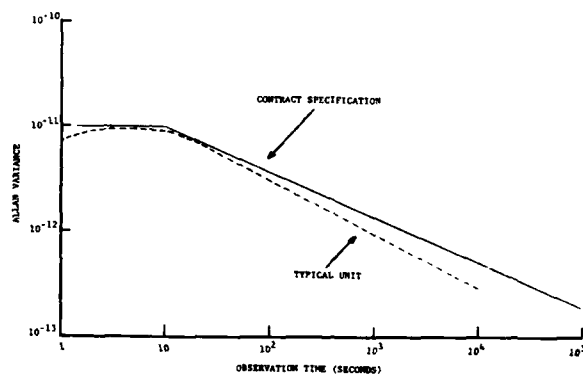


FIGURE 8: FREQUENCY STABILITY

SOME CAUSES AND CURES OF FREQUENCY INSTABILITIES (DRIFT & NOISE)
IN CESIUM BEAM FREQUENCY STANDARDS

David W. Allan, Helmut Hellwig, Stephen Jarvis, Jr., David A. Howe, R. Michael Garvey
Frequency & Time Standards Section
National Bureau of Standards
Boulder, CO 80302

SUMMARY

Frequency drift of the order of several parts in 10^{13} per year is often observed in commercial cesium beam frequency standards, and on some occasions significant changes in the white noise frequency modulation level is also observed. Recently at the National Bureau of Standards some standards with these types of problems have been analyzed and their velocity distributions measured. A comparison of the changes in drift and noise performance with measurements of the velocity distribution leads to some interesting interpretations:

Changes in focusing voltage at or in the vicinity of the detector may cause the finite surface area of the detector to act like a velocity selector; i.e., the detection efficiency of cesium atoms mapped across the surface of the detector changes with time as a result of changing electric fields which focus the cesium ion-beam. Changes in the microwave power cause changes in the most probable cesium atom velocity, which transduce via the Ramsey cavity phase shift into frequency changes.

The magnitude of the cavity phase shifts in a cavity was inferred by reversing the current through the C-field inducing mixing of the m_F states

I. INTRODUCTION

A basic limitation in keeping accurate time with a cesium beam atomic clock results from its long-term frequency instabilities. It is the purpose of this work to investigate in a non-destructive way some of the causes of these instabilities and, in some cases, propose some possible improvements.

II. BACKGROUND

There are several possible causes of frequency instabilities and drifts in cesium beam frequency standards. Specifically some of these are: microwave spectrum impurities and asymmetry; distortion of the modulation frequency; excess noise at the modulation frequency; ac and dc instabilities in the servo demodulator circuitry; servo loop gain degradation; frequency drift in the quartz crystal oscillator locked to the cesium resonance causing frequency drift in the cesium standard's output due to too low a loop gain; drift and

instabilities in the magnetic field in the C-field drift region; gradual changes in the cesium beam optics; pulling by neighboring field dependent lines; etc. Basically, the performance of the electronics is dependent on the beam tube's characteristics.

It has been shown that one of the main limitations to the accuracy of a cesium beam primary frequency standard is the cavity phase shift associated with the Ramsey microwave excitation region as seen by the cesium beam [1,2]. At this point we will develop a model which indicates that this is also one of the main limitations to the long-term stability in cesium beam frequency standards. From Ramsey [3] we have the transition probability near the Ramsey resonance for cesium atoms with a velocity, v , as follows:

$$P_{p,q} = \sin^2 2b\tau \cos^2 \frac{1}{2}(\omega_0 - \omega - \phi/T)T, \quad (1)$$

where b is proportional to the square-root of the microwave power in the cavity, τ is the time the atoms spend in each end of the Ramsey cavity, ω_0 is the natural resonance frequency of the cesium atoms, ω is the microwave interrogation frequency, ϕ is the cavity phase shift between the cavity ends, and T is the flight time of the cesium atoms between the cavity ends (of course $v = L/T$ where L is the distance between the cavity ends). Now from eq. (1) it is clear that the center of the resonance will be shifted from ω_0 by an amount ϕ/T causing a frequency shift in the standard of that amount due to the cavity phase shift. In most commercial cesium tubes, the "average" T is of the order of 1 ms, so in order to have a frequency shift of less than 1 part in 10^{12} requires ϕ to be less than 60 microradians or a synchronization of the microwave fields at the cavity ends to less than about 1 femto second (10^{-15} s). Furthermore, the amount of microwave power ($\propto b^2$) present in the cavity determines the most probable velocity of atoms to undergo a transition; hence, for typical non-zero kinds of cavity phase shifts encountered, the resulting frequency shift may be highly dependent on the microwave power.

A further complication arises because of geometrical imperfections and finite conductivity associated with the ends of the Ramsey cavity. This causes the cavity phase

shift not to be single valued for a given cavity but to have a value dependent upon where the cesium beam trans-
verses the cavity opening. In other words, the phase shift, ϕ , is a function of the coordinates across the cavity opening, which has been called the "distributed cavity phase shift" effect [4,5]. The distributed cavity phase variations may be of the same order as the basic phase difference. Frequency shifts result which are a function of the beam trajectory location which itself depends on other parameters, such as microwave power.

In practice, of course, there is a velocity distribution, $\rho(v)$, which convolves with the transition probability to give the typical kinds of Ramsey spectra observed. Knowing the velocity distribution and changes in it may lead to insights into beam trajectories, trajectory changes and effects these changes might have on the performance of a cesium clock.

III. EXPERIMENTAL RESULTS

A. Analysis of an older commercial cesium beam standard with significant frequency drift

Figure 1 shows measured frequency changes of a particular commercial cesium standard over more than 900 days [6]. One notes a positive frequency drift over the first 100+ days and then a negative drift of about -2×10^{-12} /year over the remainder of the data. One notes also a significant increase in the white noise FM level toward the end of the plot in Figure 1. For a white noise FM process the spectral density is given by $S_y(f) = h_o/y$ where y is the fractional frequency, f is the Fourier frequency, and h_o is the noise level. In this case one may write $\sigma_y(\tau) = \sqrt{h_o/2}$, where τ is the sample time in seconds for the two-sample sigma. The values of $\sqrt{h_o/2}$ went from 1.2×10^{-11} to 2.6×10^{-11} to 8×10^{-11} for 1975, mid-1976 and for 11 November 1976, respectively. The velocity distribution of this particular cesium tube was measured 6 December 1976 and is shown in Figure 2 as curve a. A modification was made per the manufacturer's recommendation to the circuit associated with the cesium detector. The velocity distribution was again measured on 14 December 1976 as is plotted as curve b in Figure 2. Note that the maximum velocity moved higher by about 10 m/s. Appropriately compensating for the magnetic field correction and measuring the frequency before and after the modification showed a frequency increase of 2.8×10^{-12} and $\sqrt{h_o/2}$ went to 7×10^{-12} (an order of magnitude improvement). The drift was measured over about a month and was

observed to be back positive at about the rate of $+8 \times 10^{-13}$ /year. One will note that the frequency also returned to about its same value as at the beginning of the life of the tube.

The low velocity secondary hump seen in curve b of Figure 2 is also characteristic of this type of tube at the beginning of its life [5]. It seems probable in this case that gradual changes in the characteristics of the detector (the only parameter changed in the modification) cause a preferential detection of different velocities of atoms, which in turn may cause a frequency drift via the cavity phase shift. We cannot offer a more detailed explanation of the actual physical changes (probably in the detector) beyond these phenomenological statements. Much more work in this area is required.

B. Analysis of a new standard

Figure 3 is a plot of the average daily frequency of the first 2 months of the life of a new cesium standard. It appears to be drifting positively at a rate of about 1×10^{-12} /year. Following this data a velocity distribution, $\rho(v)$, was taken on 17 March 1977 and is shown in Figure 4. From the $\rho(v)$ curve, one can generate [7]: the corresponding Ramsey spectrum as shown in Figure 5; the beam current intensity, I_B , at the detector as a function of microwave power as shown in Figure 6 (curve g_o), for which the frequency shift is:

$$v_b = + \frac{\phi v_p}{2\pi L} \quad (2)$$

where v_p is an appropriate average velocity.

The typical operating point for the microwave power is with I_B set to a maximum, called "optimum power". Clearly, at this point the v_p curve has a significant non-zero slope; i.e. if the microwave power changes, then v_p will change, and if there is a cavity phase difference or a distributed cavity phase shift then the frequency will change.

This particular standard has a temperature coefficient of about $-1 \times 10^{-13}/^\circ\text{C}$ in an operating environment of 20°C to 30°C . It was believed that a main cause of this was due to microwave power changes as a function of temperature; i.e., one could see a noticeable change in the beam current monitor voltage as a function of temperature as shown in Figure 7. A similar temperature coefficient was obtained by applying localized heating at the harmonic generator which generates the microwave

signal. No attempt was made to apply localized heating at the frequency multiplier which starts at 5 MHz.

Now, in the case of a mono velocity beam with the appropriate (optimum) microwave power this velocity is the same as v_p . From Eq. (1), one calculates that the offset frequency of the first side lobe is simply:

$$\nu_1 - \nu = \frac{v_p}{L}. \quad (3)$$

This equation will be somewhat modified when Eq. (1) is convolved with a velocity distribution, but we can verify that the offset frequency of the side lobe is strongly and nearly linearly dependent on v_p over the power range of interest. Also, this dependence is, by far, the predominate determinant of the frequency separation of the side lobe from the main peak. Since v_p depends on microwave power, so will ν_1 . Figure 8 is a plot of the frequency dependence of the first upper Ramsey side lobe as a function of temperature as measured at the harmonic generator.

Using the v_p curve in Figure 6 allows one to estimate the change in microwave power per degree assuming Eq. (2) is nominally valid. Doing so indicates that in this particular standard the microwave power changes about $-0.1 \text{ dB}/^\circ\text{C}$.

The fractional frequency shift of the cesium tube (main frequency peak) is given as a function of ϕ and v_p in Eq. (2). If the cavity phase shift ϕ is of the order of $\pm 0.1 \text{ m rad}$ [1,2], then for a tube with $L = 17 \text{ cm}$ a velocity change of 1 m/s (about 1%) causes $\Delta\nu/\nu = 10^{-14}$. From Figure 6, 1 m/s change is caused by a power change of approximately $P = .6 \text{ dB}$ at optimum power. In other words, flicker noise of the microwave power in the amount of $.6 \text{ dB}$ would limit the standards' performance to a flicker floor of 10^{-14} . Furthermore, smaller values of L (shorter tubes) would lead to worse performance according to Eq. (2), as well as steeper slopes of the $v_p(b)$ curve (Fig. 5), for constant ϕ and v_p . A steeper $v_p(b)$ curve is related to broader velocity distribution (v) and a narrow velocity distribution results in relatively smaller dependence of v_p on b (at optimum power).

Data from many tubes bear out this assertion; i.e., the shorter the tube and the broader (v) , the worse the flicker floor in ν_1 [3]. Therefore, with conventional tube design, the long-term stability can be fundamentally improved by going to relatively long tubes and narrow velocity distributions while keeping the

"scale" factor, (cavity phase difference and distributed cavity phase shift) as small as possible via careful fabrication and the use of small beam openings in the cavity.

Other mechanisms which can transduce microwave power changes to frequency shifts include pulling by neighboring field dependent lines and microwave impurities or assymetry. We report on a net effect; studies of other mechanisms are under way.

Based on the measurements reported here, a novel power control could be built by constructing a servo to keep the first Ramsey side lobe at a constant frequency offset by controlling the microwave power. This may significantly improve the long-term stability of such a cesium clock. For a stability performance equivalent to 10^{-15} , v_p must be controlled to about 0.01%; this may be accomplished by sensing and maintaining the frequency separation of the side lobe from the main peak to the same precision; This means to about 0.01 Hz or 1×10^{-11} . This is quite feasible and would lead to response time of such a servo of about 10s.

We reversed the magnetic field of a commercial cesium beam tube which resulted in a frequency shift of the order of 1×10^{-11} . It is unlikely that this effect is attributable to a discrepancy between H^2 (measured) and H^2 (seen by the atoms).

Consider, however, the following condition present in most commercial tubes (see Figure 9):

- (A) The region between the state selecting magnet and the shielded C-field region is characterized by a rapid change in magnetic field strength of over five orders of magnitude. Magnetic trimmers are installed to assure a smooth transition without abrupt changes and reversal of magnetic field components. This condition is optimized for the given magnetic shielding and C-field conditions. Reversal of the C-field leads to a destruction of this condition and thus to mixing of the m_F -states via Majorana transitions.
- (B) All presently used cavity structures in cesium beam tubes not only have a small cavity phase difference between the two interrogation regions of the Ramsey cavity, but also a variation of the phase across the beam opening in the cavity. This latter effect, the distributed cavity phase shift, causes

a frequency shift which depends on the location of the atomic trajectories with respect to the beam tube axis [6,7]. The magnitude of this effect, which limits the accuracy of primary standards, was determined to be as large as several parts in 10^{13} for NBS-6 [2] and is estimated to be of the order of 10^{-12} to 10^{-11} for commercial tubes.

From (A) and (B) above, we construct the following explanation for the frequency shift under C-field reversal: In the original C-field direction, negligible state-mixing takes place and the interrogated trajectories of the C-field region are those of the originally selected $m_F = 0$ state. In the reversed field configuration, mixing produces some $m_F = 0$ states in all trajectories. Since the m_F states have different [fig. 9] magnetic moments, the trajectories are different, leading via the distributed cavity phase shift and a possible change in the effective velocity distribution to a frequency change.

To summarize, because of the presence of the (distributed) cavity phase shift, the frequency of the cesium beam tube depends on the degree of mixing of m_F states in the region between state selector and C-field and is bounded by the magnitude of the distributed cavity phase shift. In order to avoid additional uncertainties in the evaluation of primary standards via beam reversal, the degree of m_F state mixing must be made identical for both transition regions on either side of the C-field region. This can be done by the techniques used in NBS-6 and many commercial tubes, i.e. by the use of trimmers, or by assuring complete mixing via properly designed magnetic field discontinuities in the transition region or via use of low frequency fields coupling the m_F -states. The former approach is sensitive to external fields; it also leads to a "correct" C-field polarity versus a "wrong" one [9]. Both approaches warrant further investigation to determine their full impact.

IV. CONCLUSIONS

It now seems clear that significant improvements are still possible in cesium beam frequency standards. The microwave cavity phase shift appears to be the transducer of many of the inaccuracies and long-term instabilities in such a standard. One needs to either reduce that shift or provide better control of those elements which transduce via this shift to cause inaccuracies and long-term instabilities. As an example a

method is proposed in this paper of improving the control of the microwave power by one or two orders of magnitude which may yield parts in 10^{15} stability. Better cesium atom detectors, which are not cross section dependent with changes in the associated parameters, could significantly improve the long-term performance.

Magnetic field related frequency shifts can occur not only via changes in the C-field but via m_F -state mixing between state selector and C-field. However, a minor redesign could control this effect.

V. ACKNOWLEDGEMENTS

The authors wish to thank the SAMSO staff (Contract SMS 70105) who have supported a good part of this work and have permitted it to be published. Don Percival and G.M.R. Winkler of USNO have been extremely cooperative with equipment and relevant data and we wish to thank them for same. We wish also to thank Howard E. Bell who has assisted greatly in helping to set up the experiments.

REFERENCES

- [1] Glaze, D. J., et al. "NBS-4 and NBS-6: The NBS Primary Frequency Standards," *Metrologia* **13**, pp. 17-28 (1977).
- [2] Wineland, D.J. et al. "Results on Limitations in Primary Cesium Standard Operation," *IEEE Trans. on I&M*; **IM-25**; No. 4, pp 453-457 (1976).
- [3] Ramsey N. F., "Molecular Beams," *Oxford University Press* (1956).
- [4] Jarvis, Stephen Jr., "Molecular Beam Tube Frequency Biases Due to Distributed Cavity Phase Variations," NBS Technical Note 660, January 1975.
- [5] H. Hellwig, et al. "Atomic Masses and Fundamental Constants 5," edited by J.H. Sanders and A.H. Wapstra, Plenum Press, p. 330, (1976).
- [6] Data supplied by Don Percival of U.S. Naval Observatory (USNO). Also see Percival, D., "Prediction Error Analysis of Atomic Frequency Standards," to be published in the Proc. 31st Annual Symposium on Frequency Control, June 1977.
- [7] Jarvis, Stephen Jr., "Determination of Velocity Distributions in Molecular Beam Frequency Standards from Measured Resonance Curves," *Metrologia* **10**, pp. 87-98 (1974).
- [8] Howe, D. A., "Velocity Distribution Measurements of Cesium Beam Tubes," *Proc. 30th Annual Symposium on Frequency Control*, pp. 451-456 (1976).
- [9] In NBS-6 as well as in NBS-5 all measurements reported in the past were done with the "correct" C-field polarity, i.e., preservation of the $m_F = 0$ state was assured by properly aligned trimmers for both beam directions. We believed until now that the trimmers helped only to improve the signal strength.

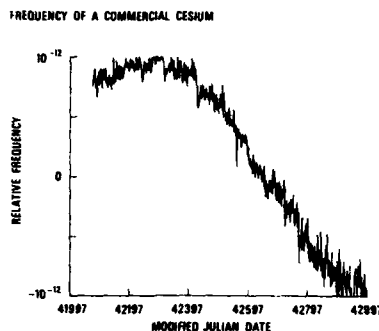


FIGURE 1 A plot of the relative frequency of a commercial cesium beam frequency standard with respect to the USNO frequency reference. The frequencies are one day averages and the plot is of about 900 days starting about the beginning of 1974. (Taken from ref. [4])

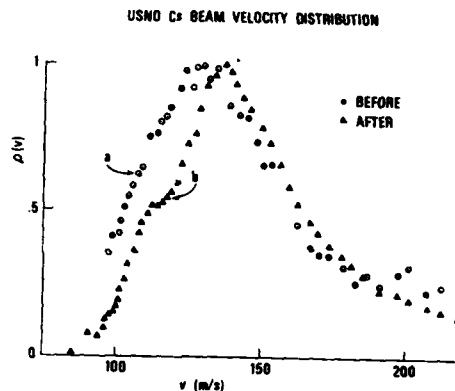


FIGURE 2 Curve a is a plot of the velocity distribution of the commercial cesium beam frequency standard. This distribution was measured (6 Dec. 76) after the data were taken at USNO (plotted in Fig. 1) and before modifying the detector. Curve b is a velocity distribution measured (14 Dec. 76) after the manufacturer's recommended modification to the cesium detection system.

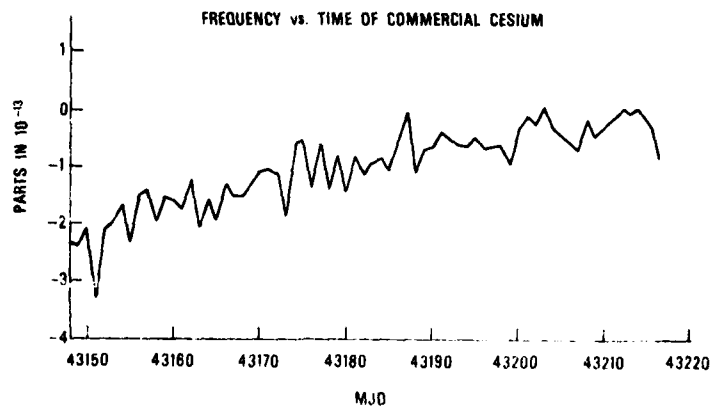


FIGURE 3 A plot of the frequency of a newly received commercial cesium beam frequency standard relative to an NBS frequency reference. Each point is a one day average, starting 5 Jan. 1977.

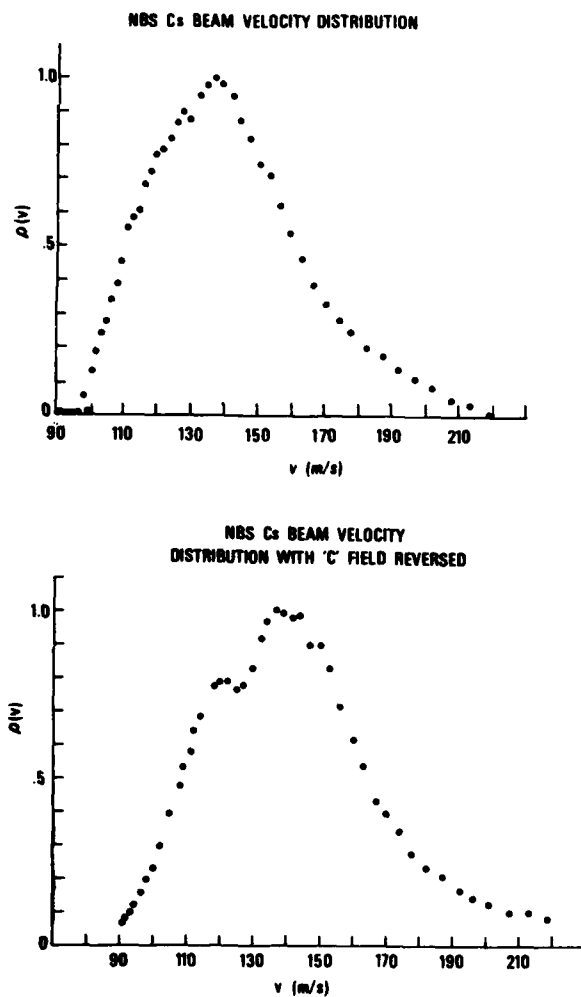


FIGURE 4 A velocity distribution measured (15 Mar 77) on the standard of Fig. 3.

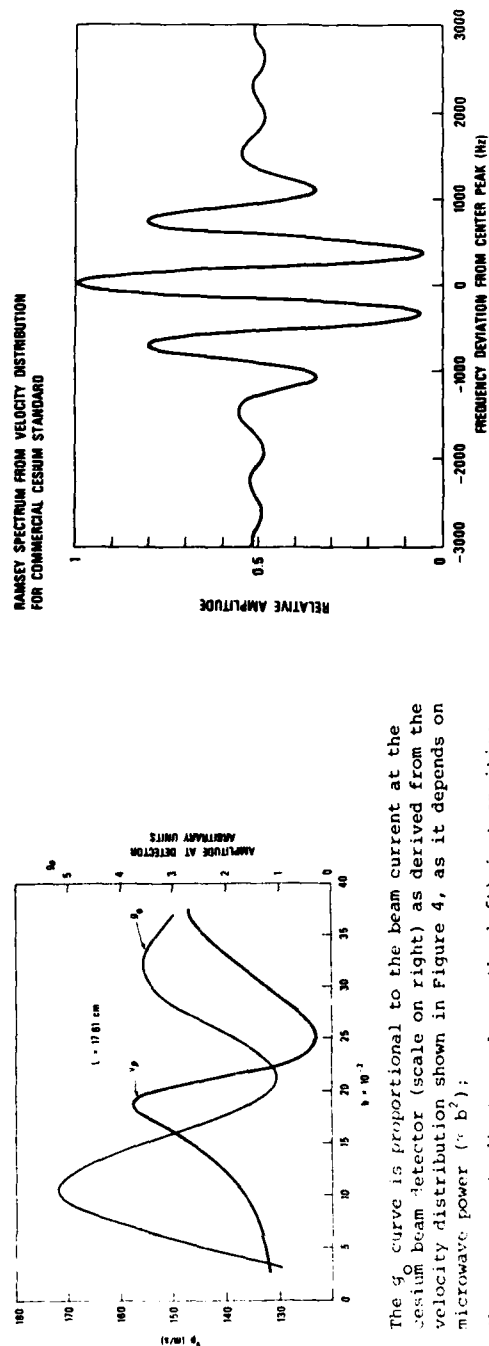


FIGURE 5 The Ramsey spectrum for a commercial cesium beam frequency standard as derived from the velocity distribution shown in Figure 4.

(a) The ρ_0 curve is proportional to the beam current at the cesium beam detector (scale on right) as derived from the velocity distribution shown in Figure 4, as it depends on microwave power ($\propto b^2$);

(b) The ρ_1 curve (ordinate scale on the left) is transition velocity as a function of microwave power as derived from the $\rho(v)$ curve in Figure 4. Both curves are for the same commercial cesium beam frequency standard.

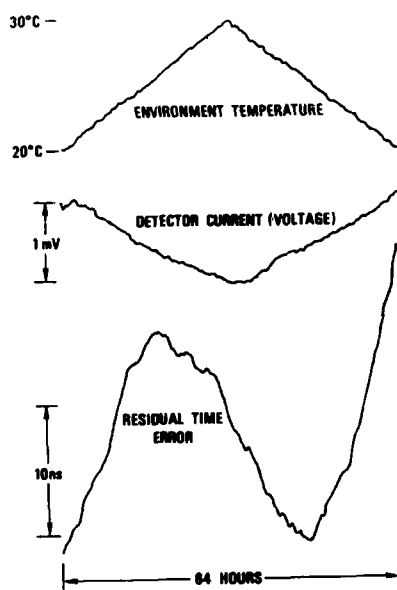


FIGURE 7. A plot of the changes in the residual time error and in the monitor voltage of the beam detector current as a function of the environmental temperature for the same commercial cesium beam frequency standard of Figures 3 and 4.

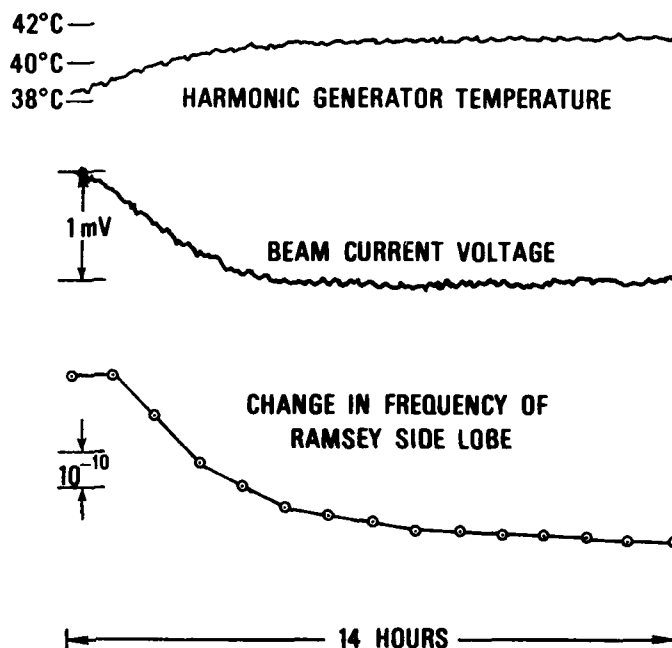


FIGURE 8. A plot of the dependence of the temperature at the harmonic generator (microwave source for the Ramsey cavity), of the monitor voltage of the beam detector current, and of the frequency of the upper first side lobe of the Ramsey spectrum as a function of environmental temperature of the same commercial cesium beam frequency standard of Figures 3 and 4. The environmental temperature was moved from about 23°C to about 27.5°C and left at the latter temperature for the last several hours of the experiment. The large dependence of the frequency of the Ramsey side lobe is obvious and was simple to measure; i.e., a frequency measurement was made every hour with an averaging time of 11 seconds.

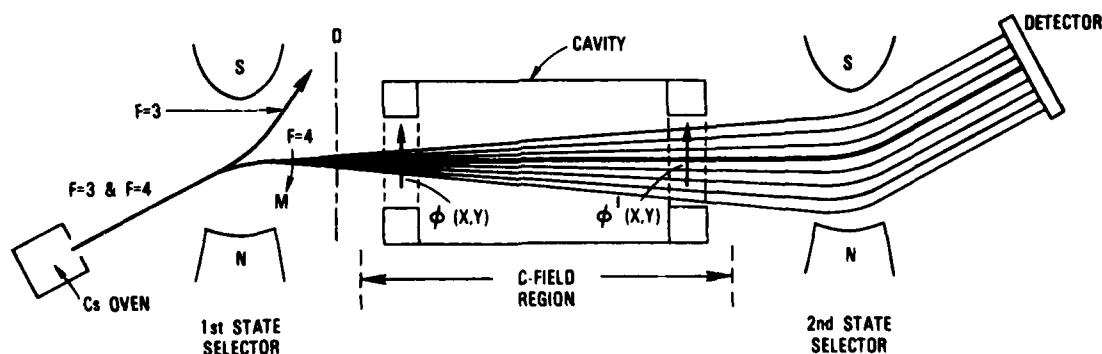


FIGURE 9. Frequency shifts due to distributed cavity phase shifts, $\delta(x,y)$ and $\delta'(x,y)$ in conjunction with m-state mixing.

D depicts an assumed magnetic field discontinuity causing more or less complete mixing of the m-states.

For illustration purposes, only one of several possible beam optics configurations is depicted.

RESULTS WITH THE SPECIAL-PURPOSE AMMONIA FREQUENCY STANDARD*

David J. Wineland, David A. Howe, and Michael B. Mohler
Frequency and Time Standards Section
National Bureau of Standards
Boulder, CO 80302
(303) 499-1000, extension 3224

SUMMARY

A special-purpose frequency standard and clock has been developed, featuring a novel combination of stability and accuracy performance, shock and temperature insensitivity, instant turn-on characteristics, and the potential for low weight, low power consumption and fabrication costs. This device should be able to fill a metrology need not satisfied by presently available atomic and quartz-crystal standards.

The device is based on the well-known 3-3 transition in ammonia (~ 23 GHz), which provides the frequency reference for a ~ 0.5 GHz oscillator. This oscillator is a novel stripline transistor oscillator of high spectral purity. Its output is multiplied in one step to K-band and the resulting output is passed through a waveguide cell containing ammonia. The detected absorption feature is used to frequency lock the 0.5 GHz oscillator to line center. To accomplish this, the oscillator is frequency modulated at ~ 10 kHz and is locked by nulling the third harmonic of the detected output. This technique discriminates to a high degree the effects of "background slope" pulling. In addition, pulling from the microwave cavity absorption cell is diminished by locking the cavity to the ammonia transition by nulling the fifth harmonic of the detected output. A fixed output frequency between 5 and 10 MHz is provided by direct division from 0.5 GHz.

The observed stability is 2×10^{-10} from 10 to 6000 sec., and reproducibility is estimated to be $\pm 2 \times 10^{-9}$. The rather broad linewidth of ammonia (~ 100 kHz due to Doppler broadening) reduces overall resolution but allows a short (~ 0.15 ms) servo attack time, thus reducing the acceleration sensitivity of the primary 0.5 GHz oscillator. In this respect the ammonia absorption device may have a unique advantage. Linewidths of most other precision oscillators are considerably narrower, resulting in longer attack times needed to lock the primary (vibration sensitive) oscillator to the (vibration insensitive) frequency reference. Vibration sensitivities as small as 10^{-11} s²/m

(10^{-10} /g) appear feasible. Servo offset problems have been substantially reduced by using digital demodulators in the feedback networks. Working at pressures where collision broadening contributes to the linewidth allows one to easily null out pressure shifts; shifts smaller than 3.8×10^{-9} /Pa (5×10^{-10} /μm) have been observed. Power consumption should be < 3 W and expected size of a working device $\sim 10^3$ cm³. With further development, improvements in performance can be expected.

I. INTRODUCTION

The special-purpose ammonia frequency standard has been developed because of the need for an oscillator satisfying specific requirements not found in other precision oscillators. The design and goals of the present project have been discussed elsewhere [1]; briefly we seek an oscillator with accuracy (i.e., long-term reproducibility) in the 10^{-9} range and stability in the 10^{-10} range. Other desirable features to be included are environmental insensitivity, (i.e., low sensitivity to acceleration, temperature, magnetic and electric fields) and instant turn-on capability. Finally it is desirable that the device have low weight and volume and small fabrication cost.

To realize such an oscillator we have investigated the scheme shown in Fig. 1. This basic scheme was used in the first "atomic" clock [2] and was revived because of its simplicity. In Fig. 1 a primary oscillator is referenced to the 3-3 transition in $N^{15}H_3$ (22.79 GHz). For simplicity the frequency of the primary oscillator is chosen to be ~ 0.5 GHz; this allows one to multiply in one step to the ammonia frequency and at the same time directly divide the oscillator output frequency to produce an output frequency near 5 MHz which is harmonically related to the ammonia frequency. (Multiplication factor ≈ 4400). The oscillator is a simple stripline oscillator, the multiplier is a waveguide mounted step recovery diode, the ammonia transition cell is a simple closed cell and the dividers are readily available I. C. modules.

Below we discuss the system and the results obtained. It is convenient to cover first the main components of the system individually: oscillator and divider, multiplier, ammonia cell and microwave cavity, and

*This research was supported by the Advanced Research Project Agency of the Department of Defense and was monitored by ARPA under Contract #D140.

servo system. We then discuss results obtained on stability, accuracy (reproducibility), ammonia cell development and sensitivity to environmental parameters. Finally, we note future possibilities based on our present results.

II. ~ 0.5 GHz PRIMARY OSCILLATOR AND DIVIDER

In reviewing the possible oscillator designs, it appeared that an oscillator using a simple LC resonator should be investigated. Advantages to this design include:

- (1) wide tunability,
- (2) continuous operation under very adverse conditions (shock, vibration),
- (3) good short-term stability,
- (4) low cost.

An oscillator was developed operating at about 0.5 GHz and having a free-running stability as shown in Fig. 2. The curves were obtained with a divider chain ($\div 100$) after the 0.5 GHz oscillator. These data were computed using the two-sample variance for different averaging times [3]. The bandwidth of the measurement system affects the variance in the case of white and flicker of phase type noise; therefore, two curves are plotted around the averaging times of interest (~ 10 ms). The oscillator features a P. C. board etched strip as a transmission line resonator (stripline resonator). In the design of a high-performance stripline oscillator, we must address three principal problems [4]:

- (1) minimization of resonator losses,
- (2) minimization of additive transistor noise, and
- (3) shock and vibration isolation of the resonator.

There are other problems which must be looked at, but these three represent the major contributors to degradation in stability.

Radiative loss is minimized by adopting a three-layer sandwich etch technique. In this design, two ground planes are used on the top and bottom surfaces of the P. C. board with the stripline centered in the dielectric. To keep loss to a minimum, fiberglass-*teflon* which has a small loss tangent of about 10^{-3} is used for the dielectric. The stripline itself is a 7 cm length of copper which is 1 cm wide and 2 mm thick. Contact resistance is minimized by using silver-solder on all connections. The unloaded Q of the resonator at 0.5 GHz is about 400. Loaded Q of the resonator is maximized by the use of a field-effect transistor as the active element [5]. It is chosen to have a high

forward transconductance and a high cut-off frequency.

Oscillator noise is generally characterized by low frequency (near carrier) flicker noise behavior and high frequency (far from carrier) white phase noise. Sources of noise may be multiplicative or additive. Causes of flicker behavior are difficult to identify in many instances, but helpful in the reduction of flicker noise is the selection of a transistor which is manufactured with care and in a clean environment, since flicker noise may relate to sporadic conductance through the device due to impurities. White phase noise is usually associated with additive thermal noise of components and/or transistor parameters. One can resort to devices capable of higher signal levels in order to get above a fixed noise level. A trade-off exists between white phase and flicker noise, however, since higher device currents usually aggravate the flicker noise problem. We have arrived at a compromise solution which gives suitable performance in the NH_3 standard. The curves shown in Fig. 1 represent a higher device drive level than is common in, say, quartz-crystal oscillators.

At frequencies around 0.5 GHz, transistor package parameters (inductance and capacitance) and stray parasitic elements such as connecting lead inductance and stray capacitance all contribute to the fundamental resonance. If one is to achieve a relative frequency stability approaching 1×10^{-9} , then it is imperative to maintain resonator inductance and capacitance values stable to this level. The greatest deterrent to maintaining high inductive and capacitive stability is vibration sensitivity of the oscillator. This problem of microphonics has been reduced by using the three-layer P. C. board and rigidly mounting all components and leads with a low-loss doping compound. The oscillator is in turn rigidly fixed to an aluminum block which acts as the shield for the components. The test block weighs about 3 kg. Depending on the application, one can rigidly mount or soft mount the oscillator into a system. If rigidly mounted, structure-born vibration is directly applied to the oscillator. A soft mount designed to isolate the oscillator from vibration can reduce the transmitted vibration at higher frequencies at the cost of increasing the vibration sensitivity at a lower frequency. Damping material can also be used to alter the vibration response.

In the ammonia standard, the problem of vibration sensitivity of the fundamental oscillator is significant in cases of shock and vibration where the frequency of the servo system is exceeded or the period of the

AD-A088 221

ARMY ELECTRONICS COMMAND FORT MONMOUTH NJ
PROCEEDINGS OF THE ANNUAL FREQUENCY CONTROL SYMPOSIUM (31ST) HE--ETC(U)
1977

F/G 9/5

UNCLASSIFIED

NL

7 of 7
Page 10



END
DATE
FILMED
9-80
DTIC

A
882

vibration is shorter than the servo attack time. A key incentive for choosing a wide spectral line is that the servo attack time can be small. Since the NH_3 resonance is ~ 100 kHz, this enables the use of fast loop response. The design of the oscillator mount should yield a vibration response in which frequencies of ~ 100 kHz and above are suitably attenuated.

The division from 500 MHz to 5 MHz uses one ECL decade divider followed by a TTL decade divider (from 50 MHz to 5 MHz). A level translator was used between these I.C.'s. The ECL divider contained an internal wideband amplifier thus allowing good isolation between divider logic and the source. In short term (< 10 ms), the white phase noise of the dividers set a limit on the stability at 5 MHz.

III. STEP RECOVERY DIODE MULTIPLIER

It is desirable to make a multiplier module with fairly low output Q ($Q \leq 10$) and output power > 10 μW . We use state-of-the-art step-recovery diodes in a waveguide multiplier module (See Fig. 3.). In simplest terms, the problem is one of impedance matching for both the input and output circuit. For example, for the input circuit (~ 0.5 GHz) the dynamic diode impedance is $Z \approx 1\Omega$. Therefore, two π section transformers are cascaded to match to the 50Ω output impedance of the 0.5 GHz amplifier. Approximately 0.5 W input power is needed to "snap" the diode properly. To accomplish this, a microstrip hybrid class "C" amplifier is used. The amplifier, microstrip matching circuit and multiplier module are integrated into one package in order to avoid instabilities due to connections. The output circuit is composed of a shorting stub and iris coupling to form a cavity ($Q \approx 10$) with the diode approximately matched to the characteristic impedance of the narrow height waveguide. In the interest of rigidity and simplicity, shims are used rather than movable plungers. With ~ 0.6 W input power to the diode, output power of approximately 100 μW was obtained.

IV. AMMONIA GAS ABSORPTION CELL

The advantages of using ammonia as the reference "atom" are:

(1) The microwave transition of interest provides an absorption signal which is orders of magnitude stronger than those of other interesting molecules or atoms [6]. This is a significant advantage because it means that the desired signal-to-noise is obtained without resorting to impractically large microwave cell

sizes as would be necessary for almost any other gas cell.

(2) Since ammonia remains in the gas phase for the temperature range of interest (-40° to $+60^\circ\text{C}$) the device has instant turn-on capability. One must, however, note the existence of a pressure (therefore temperature) dependent frequency shift; this is discussed more fully below.

(3) The frequency of the ammonia transition is fundamental in nature and therefore essentially eliminates the need for calibration of the device.

(4) The ammonia transition linewidth is fairly broad ($\sim 100\text{kHz}$). This is a disadvantage in terms of the ultimate accuracy obtainable, but it allows the primary oscillator to be locked to the ammonia reference in very short times (> 1.5 μs). This allows a significant reduction in the acceleration sensitivity of the primary oscillator.

(a) Choice of 3-3 Transition in N^{15}H_3

The reference transition was chosen to be the 3-3 line in ammonia because of its large signal strength [6]. The N^{15}H_3 isotope was chosen because the N^{15} nucleus has no quadrupole moment and therefore it is free of the quadrupole structure which makes the 3-3 line of the N^{14}H_3 isotope asymmetric. This asymmetry causes the apparent line center to depend on FM modulation amplitude and microwave power. Uncertainties as large as approximately 4×10^{-8} could be expected from this asymmetry.

The results of three high resolution determinations of the 3-3, N^{15}H_3 frequency are:

$$\nu_o = 22\,789\,421\,731 \pm 1 \text{ Hz} \quad [7]$$

$$\nu_o = 22\,789\,421\,701 \pm 1 \text{ Hz} \quad [8]$$

$$\nu_o = 22\,789\,421\,672 \pm 55 \text{ Hz} \quad [9].$$

These measurements were taken with ammonia beam masers and show that although rather high resolution can be obtained, a conservative uncertainty in one's knowledge of the frequency is about 10^{-9} . For our work we therefore assume:

$$\nu_o = 22\,789\,421\,700 (1 \pm 10^{-9}) \text{ Hz}.$$

(b) Ammonia Containment

Most aspects of the present device could be tested in an experimental apparatus using a gas flow system for the ammonia [1]. A simple gas flow system could be used in the final device but the added complexity is a disadvantage. Therefore, it is desirable to develop permanent closed cells.

The problem of ammonia containment is well known; in a simple cell using brass or copper waveguides, reaction of ammonia with the cell walls causes rapid disappearance of the signal particularly if water vapor is present or the waveguide walls are not sufficiently clean. Also, even in a clean environment, chemisorption with certain metals is noted [10]. Various inert cell coatings were considered [1], however it was felt that a simpler approach would be to make cells from inert materials whose cleanliness could be insured.

Two basic problems are encountered in obtaining such a cell: (1) a provision must be made for evacuating, baking, and cleanly sealing the cell, (2) the cell must form part of or be included in a microwave cavity for interrogation by the microwave radiation. This implies that the material used for the windows must be nearly lossless and be easily bonded to the rest of the cavity. Two basic schemes were tried and although not enough work has been completed on cell development, the results are encouraging.

First cells were made with quartz cylinders of rectangular cross section which could be inserted into X-band waveguide. Both ends were extended and drawn into points to provide a seal. Quartz was used because its microwave losses are negligible. The cells were baked under vacuum at 300°C for 30 hours at a pressure of $< 10^{-4}$ Pa (1 Torr \approx 133 Pa). Ammonia was then backfilled at pressures between 0.13 and 0.67 Pa (10^{-3} Torr to 5×10^{-3} Torr). No attempt was made to purify the ammonia which was contained in a lecture bottle and admitted to the cell through a metal leak valve. The cells were then sealed off by heating a quartz pinch-off.

A second type of cell was made of stainless steel K-band waveguide. Windows were made of quartz which were mated to brass flanges with pressed indium seals. Ammonia was admitted through a copper pinch-off which was silver soldered to the broad face of the waveguide in which small holes were drilled. The cells were baked at 100°C for 50 hours at a pressure of $< 10^{-4}$ Pa. The gas was admitted as above (without purification)

and the copper pinch-off was then sealed.

(c) Microwave Cavity for Ammonia

Conceptually, the simplest approach would be to pass the microwaves through the ammonia and servo the frequency of the oscillator to the point of maximum absorption. No errors would occur with this method if there were no reflections at the ammonia cell interface and if the source and detector were perfectly matched to the microwave guide. Such is hardly the case in practice and since reflections occur at both sides of the ammonia cell, it is effectively contained in a cavity.

Frequency pulling due to cavity mistuning is a familiar problem to designers of atomic clocks. In a passive standard such as the one discussed here, the frequency at which microwave absorption is maximum is given by [11]:

$$(\nu - \nu_0) = \frac{Q_c}{Q_l} (\nu_c - \nu_0) \quad (1)$$

where ν_0 = unperturbed ammonia frequency
 ν_c = cavity center frequency
 Q_c = microwave cavity Q
 Q_l = ammonia transition Q

and the expression is valid when: $(\nu_c - \nu_0)/\nu_0 \ll 1/Q_c$, $Q_l \gg Q_c$.

This result follows simply from the fact that the ammonia and microwave cavity form a system of coupled complete oscillators. Therefore varying the frequency of the cavity changes the apparent frequency of the ammonia line. In a cavity formed from a section of waveguide the importance of reflections is illustrated by example: If we made an ammonia cell from a section of copper WR-42 waveguide of length l and if the windows had (real) voltage reflection coefficients of value Γ_v , then the effective Q of the resulting cavity is given by:

$$Q_c \approx 4l \lambda_g \Gamma_v / \lambda_0^2 \quad (2)$$

where λ_g = guide wavelength
 λ_0 = free space wavelength.

If we choose $\Gamma_v = 0.2$, $l \approx 50$ cm [12] and since $\lambda_g \approx 1.59$ cm, $\lambda_0 = 1.31$ cm at the frequency of interest then $Q_c \approx 115$. From Eq. 1, we see that we would have to tune the cavity to 0.02% of its linewidth to obtain 10^{-9}

accuracy in the output frequency. Since the expansion coefficient for copper is $\alpha_c = 1.5 \times 10^{-5}/^\circ\text{C}$ and $Q_L \approx 2 \times 10^5$ then the frequency dependence on temperature due to cavity pulling would be:

$$\frac{\partial \nu}{\partial T} = \frac{Q_c}{Q_L} \alpha \approx 10^{-8}/^\circ\text{C}.$$

Because of this rather strong sensitivity and because it is difficult to make reflectionless windows for the cavity, one must servo the cavity to line center. A few possibilities exist for accomplishing this. One is to separately sense the cavity frequency with microwave power applied symmetrically to either side of the cavity resonance [13]. However, frequency-to-amplitude conversion is a severe problem with the low cavity Q obtained here and such a method is precluded. If Q_L could be varied, then one could look for zero change in output frequency when this is done; this assures $\nu_c - \nu_o = 0$ in Eq. 1. This could be accomplished by symmetrically broadening the line by applying a magnetic field [14]. Unfortunately this broadening is only ~ 7 MHz/T (1 Gauss = 10^{-4} T) and therefore a rather large magnetic field modulation is required to appreciably broaden the line. Another disadvantage of this scheme is that it requires a second reference oscillator in order to detect frequency changes when the magnetic field is changed.

A third method exists and, to the authors' knowledge, has not been used previously. It is discussed in the next section.

V. SERVO SYSTEM

The basic requirement of the servo system is that it force the frequency of the primary oscillator (~ 0.5 GHz) to be at a subharmonic of the ammonia transition frequency. Of course, various systematic effects shift the apparent frequency of the ammonia transition; these must either be eliminated or controlled in a known way. Although the performance of the device is not high when compared to a state-of-the-art atomic clock, the demands on the servo system are rather high. This is because we are trying to resolve the rather broad resonance feature (i.e., "split the line") to about 10^{-5} or 0.001 percent. Therefore, state-of-the-art techniques must be employed. Below we discuss the effects of cavity pulling, distortions in the source and detector and servo offsets and drifts.

(a) Cavity Pulling

We assume here that the microwave source and detector are perfectly flat, that is, the power output

from the source and the detected power do not depend on frequency. Eq. 1 yields the frequency ν where the absorption is a maximum. To facilitate the detection of this condition, source frequency modulation is used. This technique is well known and is used in other atomic clocks including cesium, rubidium, and the passive hydrogen standard [15]. Basically, we frequency modulate the source at frequency ν_m so that the time dependence of the microwave field from the source is:

$$E = E_o \cos(\omega_s t + \frac{\Delta\omega}{\omega_m} \cos \omega_m t + \phi),$$

where ω_s = average frequency of source,

$\Delta\omega$ = peak frequency excursion,

$\omega_m = 2\pi\nu_m$ (angular modulation frequency),

ϕ = arbitrary phase angle,

$\Delta\omega/\omega_m \equiv$ modulation index.

When $\nu_s - \nu_o \ll \Delta\nu_L$ ($\Delta\nu_L$ = ammonia linewidth) the detected signal will have an oscillating component at frequency ν_m ; the phase and amplitude of this component will vary as a function of frequency as shown in Fig. 4(a). These curves are the voltages observed at the outputs of the mixers in Fig. 5. This "dispersion" curve is used to servo the primary oscillator to the apparent line center by finding the condition at which the dispersion curve has zero output. When $\Delta\omega/\omega_m \ll 1$, the frequency at which this occurred is given by Eq. 1. However for $\Delta\omega/\omega_m \approx 1$, Eq. 1 must be slightly modified. Also we note the appearance of higher harmonics of ν_m in the detected output; in particular, the amplitudes and phases of the higher odd harmonics have similar "dispersion" characteristics. Figures 4(b) and (c) show the dispersion curves for the third and fifth harmonics respectively when $\Delta\omega/\omega_m$ is adjusted to give maximum slope near the center of the pattern. (Vertical scales are changed for each part of Fig. 4.) We can then use the dispersion curves of the higher harmonics to lock to line center. Two important points should be made:

(1) For large modulation amplitudes $\Delta\omega/\omega_m \gg 1$ it can be shown that Eq. 1 is modified to the form:

$$\nu - \nu_o = K(n) \frac{Q_c}{Q_L} (\nu_c - \nu_o) \quad (3)$$

where $K(n)$ is close to unity but varies by factors of two or three as $\Delta\omega$ is varied. It is also a function of the harmonic number observed (n) and in general $K(n) \neq K(n')$ for the same $\Delta\omega$. We have used this last fact to simultaneously servo the oscillator and cavity to line center therefore avoiding cavity pulling. As shown in

Fig. 5, the third harmonic dispersion is used to lock the multiplied primary oscillator to apparent line center, i.e., the condition:

$$\nu - \nu_0 = K(3) \frac{Q_c}{Q_l} (\nu_c - \nu_0) \quad (4)$$

is satisfied. The fifth harmonic dispersion is used to lock the cavity to the line center, i.e., the condition:

$$\nu - \nu_0 = K(5) \frac{Q_c}{Q_l} (\nu_c - \nu_0) \quad (5)$$

is satisfied. Since $K(3) \neq K(5)$, conditions (4) and (5) can be satisfied simultaneously only if $\nu_c - \nu_0 = \nu - \nu_0 = 0$.

(2) The use of the above scheme causes some loss of signal strength; to give an idea of this loss we have measured the ratio of the slopes $[SL(n)]$ of the dispersion curves when the microwave power is kept fixed and when the slope is maximized for each harmonic.

We measured:

$$SL(1) : SL(3) : SL(5) = 1.0 : 0.42 : 0.26$$

From the standpoint of signal strength it would be better to use the fundamental and third harmonic locks in the above scheme; however, conditions (outlined later) led us to choose the third and fifth harmonic locks. We note that a further slight reduction in signal strength or slope ($<10\%$) is observed when the system is optimized for the third and fifth harmonic locks simultaneously.

The microwave cavity is formed by a combination of shunt impedances due to the windows and added reactances at both ends of the ammonia cell. Electronic tuning is provided by a varactor diode at one end of the cell.

(b) Distortions in the microwave source and detector

In general, the source and detector are not "flat" with frequency; that is, frequency to amplitude conversion occurs which shifts the apparent frequency of the ammonia transition. The most serious problem occurs in the source. Briefly, frequency-to-amplitude conversion occurs because of the (slightly mistuned) reactances in the source. In time domain, a qualitative picture is given by assuming that as the frequency of the source swings below ν_0 , its amplitude increases; as its frequency swings above ν_0 , its amplitude decreases. Assuming that $\nu_c = \nu_0$, and $\nu = \nu_0$, then the signal from the ammonia absorption would be stronger on the low side of ν_0 than on the high side. Equivalently, there exists a residual signal component at ν_m (and $3\nu_m$ and $5\nu_m$) on

the detector which the servo then compensates for by shifting ν_s to a value below ν_0 . Since $\Delta\omega/2\pi \approx \Delta\nu_1$ in practice, then if AM of amplitude β exists as discussed above, we would have:

$$(\nu_s - \nu_0) \approx -\beta\Delta\nu_1 \quad (6)$$

in the locked condition. This pulling is accentuated for the fundamental dispersion lock (at ν_m) if the signal from the ammonia is only a small fraction (γ) of the signal reaching the detector because essentially all of the AM on the input rf reaches the detector even in the absence of ammonia. In this case:

$$(\nu_s - \nu_0) \approx -\frac{\beta}{\gamma} (\Delta\nu_1) \text{ (fundamental lock).}$$

However, for the third and fifth harmonic locks, Eq. 6 applies to a high degree. The third and fifth harmonics appear at the detector only because of the presence of ammonia. This is one reason for using the higher harmonic locks.

If the source is "flat" with frequency and the detector is not, the problem is not as severe as it is for the source. Since we are locking to the third and fifth harmonics of ν_m then spurious third or fifth harmonic signal is generated only because the curvature (nonlinearity) of the detector converts the FM (at frequency ν_m) into AM at frequency $3\nu_m$ or $5\nu_m$. Since the curvature of the detector should be small this type of offset should be negligible. This is another reason for using the third and fifth harmonic locks as noted in Sect. V(a). Third harmonic lock of the primary oscillator [16] also discriminates against background slopes (for example, due to overlapping transitions) but this is not a problem for the ammonia device.

In general, it is easier to make the detector flat and control its flatness for the source. Therefore, we have servoed out the AM (at frequency ν_m) in the source by nulling the signal (at ν_m) observed at the detector. Since the detector may not be perfectly flat, this introduces a systematic offset of the type in Eq. (1). It is important that this residual slope in the detector remain fixed in order to assure long-term stability and reproducibility. As noted below in Sect. VI(C), it may be desirable to make a detector with a specific slope which can be used to compensate the frequency shift due to pressure.

It is critically important to null the fundamental frequency signal component (ν_m) at the detector for

another reason. If this is not done, then signal at v_m can mix in the detector with the rather strong signals at $2v_m$ and $4v_m$ to give signals at $3v_m$ and $5v_m$ which gives further offsets. These offsets are avoided by exactly nulling the fundamental signal component of the detected output.

Finally, one must insure that FM distortion does not occur at the oscillator. For example, second harmonic FM distortion [17] due to a signal component of frequency $2v_m$ at the FM input causes a signal component at $3v_m$ at the detector because of the strong nonlinearity of the ammonia line. For the ammonia device, such distortion is avoided by insuring that the second harmonic of the FM input is ~ 75 dB down from the fundamental. This is easily accomplished with passive filtering.

(c) Servo phase comparators and integrators

Conventional analog phase comparators have output voltage offsets due to an asymmetry which may exist in the signal switching portion of the device. Here the signal from the microwave detector undergoes a 180° phase reversal at each half-cycle of the reference signal (third or fifth harmonic of modulation frequency). The signal path through the comparator for one half-cycle versus the other half-cycle must be identical to realize zero offset; however, this is difficult to achieve and therefore offsets will exist. These offsets not only affect the accuracy of the locked oscillator but also the stability, since they are observed to change in time.

The phase comparator is followed by an integrator to realize a second-order loop filter. Analog integrators suffer from input voltage offsets. Furthermore, the common analog integrators, having a capacitor in the negative feedback path, have finite DC gain set by the amplifier; a practical limit is about 140 dB. Capacitor leakage also degrades the analog integrator's performance.

A new phase comparator and integrator has been designed and built which employs digital electronics and can directly replace currently used analog circuitry. Since the line-splitting goal in the ammonia standard is high ($\sim 1 \times 10^{-5}$) to achieve 10^{-10} stability and the offset must be $\sim 1 \times 10^{-4}$ or lower, there was incentive to pursue techniques other than analog. Virtues to the digital system are:

1. negligible voltage offset
2. no capacitor, hence no leakage
3. infinite gain at D.C.

4. excellent low pass filter characteristics
5. excellent environmental immunity.

These positive aspects were weighted against the following observed shortcomings:

1. lower useful modulation rates necessary
2. quantization noise (additive white noise).

At modulation frequencies approaching a few kilohertz, the analog comparator and integrator begins to outperform the digital approach with regard to usable feedback gain and additive noise. The ultimate servo for the ammonia device (which can use modulation frequencies approaching 100 kHz) is one using both analog and digital techniques: the analog portion to respond in short-term and digital portion to respond in long-term.

VI. RESULTS

(a) Stability

A plot of the frequency stability obtained with the ammonia based oscillator is shown in Fig. 2. These data were computed using the two sample variance for different averaging times [3]; frequency drift has not been removed. The results using the free running ~ 0.5 GHz stripline oscillator have been discussed in Sect. II above. The data shown for 10^{-2} sec to 10 sec were taken with the oscillator locked to the apparent line center using third harmonic ($3v_m$) dispersion lock discussed in Sect. V; the cavity was unlocked for this data. The longer term data (10 sec to 6×10^3 sec) were taken with the complete system shown in Fig. 5 using a cell volume of 230 cm^3 to increase signal strength. In both cases data were taken with a gas flow system in order to directly monitor pressure; the problems of stability and cell-cavity design have been separated to simplify development.

The shorter term data were taken with a small cell (25 cm^3) to illustrate the relatively good stability obtained with small cell sizes. When the short term data (locking only the oscillator) were taken with the larger cell, (230 cm^3) approximately a factor of five improvement in stability was observed between 10^{-2} sec. and 1 sec. The cause of the flicker behavior (flattening of the $\sigma_y(\tau)$ curve) is not understood at this time but improvements could be expected, thus improving the stability by at least a factor of five over the whole range.

(b) Accuracy

For the reasons outlined in Sect. VI(c) below,

accuracy obtained will depend on system parameters such as ammonia pressure and detector slope. Since one is really interested in frequency reproducibility between units and over long periods of time (years), the reproducibility and stability of these system parameters is of primary importance. We estimate that the accuracy obtained in the above sense is approximately $\pm 2 \times 10^{-9}$ if the detector slope can be held to $\pm 2\%$ of its initially set value. This estimate is explained below; data is still needed in very long term and between different systems.

(c) Systematic frequency offsets

It is felt that the two most important systematic frequency shifts are those due to pressure and detector slope. Assuming the detected fundamental (signal at ν_m) has been electronically nulled, then there can still be a systematic offset if the detector is not flat; this offset is expressed by Eq. (6) where β is the AM of the rf and where we have assumed that the total rf signal is much larger than that due to the ammonia ($\gamma \ll 1$).

[This is the case except when the cell is very long (~10 m) and/or the reflection coefficients at the cell cavity interfaces are very high.] $\Delta\nu_1$ is written approximately as [6,14,18]:

$$\Delta\nu_1 = \left[(\Delta\nu_o)^2 + (\Delta\nu_p)^2 \right]^{1/2}$$

where $\Delta\nu_o$ (~100 kHz) is the line broadening due to all causes (primarily Doppler effect) except for ammonia-ammonia collisions, and $\Delta\nu_p$ is the linewidth due to collisions (pressure).

The frequency shift due to pressure is written as:

$$\delta\nu_p = \alpha \Delta\nu_p$$

where $\alpha \approx 0.01$ [19]. Therefore, the total shift due to detector slope and pressure is given by:

$$\delta\nu = \beta \left[(\Delta\nu_o)^2 + (\Delta\nu_p)^2 \right]^{1/2} + \alpha \Delta\nu_p.$$

The important point to note is that at high pressure ($\Delta\nu_p \gg \Delta\nu_o$) we have:

$$\frac{\partial(\delta\nu)}{\partial\Delta\nu_p} \approx \beta + \alpha$$

Therefore, if $\beta \approx \alpha$, then the change in frequency with change in pressure is very small, therefore reducing pressure (and temperature) shifts. For this condition, when $\Delta\nu_p \ll \Delta\nu_o$ then $\partial(\delta\nu)/\partial\Delta\nu_p \approx \alpha$. Experimental verification of this is given in Fig. 6 where we have

plotted fractional frequency offset (from an arbitrary reference point) versus pressure. We note that at high pressure ($P > 0.2$ Pa) that a factor of 2 change in pressure gives a fractional change in frequency of less than 10^{-9} .

Therefore, one could hope to set the detector slope such that $\beta \approx -\alpha$ and greatly reduce pressure (and temperature) sensitivity. If β changes, however, then the frequency would change so that at high pressure ($\Delta\nu_p \gg \Delta\nu_o$):

$$\frac{\partial(\delta\nu/\nu_o)}{\partial E/S} \approx \frac{\beta}{\nu_o} P$$

If $\Delta\nu_p \approx 230$ kHz, $\beta \approx \alpha = 0.01$, then if β changed by 1%, the output frequency would change by $\sim 10^{-9}$. More work is needed to improve the above scheme and of course other pressure compensation schemes are possible.

(d) Ammonia containment

Work with closed ammonia cells needs further development but first results are encouraging. For both types of cells the pressure appeared to stabilize about two days after sealing the cell. (Observed signal dropped ~30% initially.) First results indicate the stainless steel waveguide cells to be slightly better and after initial stabilization signal degradation was less than 10% after a few weeks. Future work is needed to insure cleanliness and integrity of the cells and better results are anticipated (See Sect. VI.f.(2) below).

(e) Integrator-demodulator stability

The analog and digital demodulators can both be set to give initial offsets corresponding to frequency inaccuracies of less than 10^{-10} . Therefore it is critical that their drift be minimal. To check this, the system was first locked using the digital servoes. The input signal was then also applied to the inputs of separate digital and analog demodulators and the outputs were monitored in time. If no drifts exist in these separate demodulators the output should remain constant in time. Fig. 7 shows the results of a typical measurement of this kind. The analog demodulator was the better of two commercially available high performance lock-in amplifiers. Over about a 20 minute period, the digital offset drift was smaller by more than an order of magnitude over the analog offset drift. The vertical scale in Fig. 7 is calibrated in terms of the equivalent fractional frequency offset of the primary oscillator; it therefore appears that the digital system is adequate.

Efforts continue however to find better analog demodulators.

(f) Environmental sensitivity

(1) Vibration sensitivity

Because of the fast servo attack time allowed in the ammonia standard, sensitivity to vibration should be low. This is true to the extent that the apparent absorption line is stable and that there is sufficient servo gain at the vibration frequency.

The effect of vibration suppression can be seen by observing the spectral density of phase fluctuations $S_{\phi}(f)$ of the 500 MHz oscillator under vibration. An example of such a measurement is shown in Fig. 8; one curve has the oscillator locked to the ammonia and the other with it unlocked. Vibration frequency is 40 Hz sine with a peak acceleration of 5m/s^2 . We see at least 45 dB reduction of the power in the 40 Hz additive phase spectral component when the oscillator is locked. This represents an acceleration sensitivity of at most $5 \times 10^{-11} \text{ s}^2/\text{m}$ ($5 \times 10^{-10}/\text{g}$).

In a systems design, it is desirable to build and mount the oscillator so that vibration sensitivity is as low as possible for vibration frequencies outside the servo loop bandwidth. Since the bandwidth can be made wide in the case of the NH_3 standard (approaching 100 kHz), some flexibility exists in the choice of the oscillator's mechanical design and supporting structures (see Sect. II.).

(2) Temperature sensitivity

Temperature will affect the most critical parameters affecting accuracy and long-term stability; i.e., pressure and detector slope (see Sect. VI.c.). Data was not taken on detector slope; however, to obtain $\pm 2 \times 10^{-9}$ accuracy, the slope of the detector must be maintained to about 2% over the operating temperature of the device.

Pressure sensitivity to temperature was estimated by measuring signal strength as a function of temperature on the sealed stainless steel waveguide. In this way we observed a sensitivity of $1/P \partial P/\partial T \approx$

$0.05/^\circ\text{C}$. That is, a change of 5% in pressure was observed for a change in temperature of 1°C . If the compensation scheme of Sect. VI.c is used, this gives a temperature coefficient of $\approx 5 \times 10^{-11}/^\circ\text{C}$. However, for good accuracy and stability at low ambient temperatures, some minimal temperature servoing may be required. For example, one could servo the temperature to always give a fixed signal by observing the 2nd harmonic signal at

the detector.

(3) Electric and magnetic field sensitivity

Electric fields are only of importance in the construction details of the gas cell where thermoelectric and contact potential problems may be present. A worst estimate can be based on the most sensitive hyperfine component of the (3-3) line; for this we have a relative shift of about $10^{-9} E^2$ (E in V/cm). Since electric fields surely can be limited to less than 0.1 V/cm , this does not appear to be a problem.

To first order, application of magnetic fields causes only a broadening of the ammonia line. This broadening has been observed to be [14]:

$$\frac{\partial(\delta\nu_o)}{\partial H} = 7 \times 10^3 \text{ kHz/T} \quad (1\text{T} = 10^4 \text{ Gauss})$$

If the pressure shift compensation scheme of Sect. VIC is used, then a residual frequency shift due to magnetic field will exist:

$$\begin{aligned} \frac{1}{\nu_o} \frac{\partial(\delta\nu)}{\partial H} &\approx \frac{\partial(\delta\nu)}{\partial(\Delta\nu_o)} \frac{\partial(\Delta\nu_o)}{\partial H} = 3 \times 10^{-4} \text{ a/T} \\ &= 3 \times 10^{-6}/\text{T} = 3 \times 10^{-10}/\text{G}. \end{aligned}$$

Therefore in extreme magnetic field environments some simple magnetic shielding may be required. The second order Zeeman effect is extremely small; the relative shift is about $2 \times 10^{-7} H^2$ (H in Tesla) and is therefore negligible [8].

VII. FUTURE PROJECTIONS

At the present time a fully integrated prototype has not been constructed; for simplicity the separate aspects of the device were investigated individually. However, some estimates of physical parameters can be made based on the present results:

- (a) Size requirements The lower limit on size will primarily be limited by the size of the ammonia cell. It is expected that the cell should occupy no more than 1 liter volume; hence the overall package may be from 1-2 liters in volume.
- (b) Weight requirement With proper choice of materials the expected final package weight should be less than 2 kg. For operation in extreme magnetic fields, shielding may have to be included; this will increase weight by approximately 1/2 kg.
- (c) Power requirements The basic electronic components of the present configuration are shown in Fig. 4.

The power requirements for specific portions at present:

- | | |
|---|-----------------------|
| (1) 0.5 GHz oscillator, 0.5 GHz amplifier with multiplier | 5.5 W |
| (2) divider chain ($\div 100$) | 1.0 W |
| (3) detector amplifier and servoes | <u>1.0 W</u>
7.5 W |

The 0.5 GHz power amplifier is the major drain on the power supply. Since the multiplier step-recovery diode needs about 3/4 W input, and the efficiency of the amplifier could be ~60%, improvements are expected. In an actual system we could expect total power requirements to be less than half of their present value or 3 W.

VIII. SUMMARY

Although the results obtained with the present device are encouraging, one could hope for even better results with future development. For example, the vibration sensitivity of the device was studied only very briefly but initial results show potential for unsurpassed acceleration insensitivity. Similarly, other aspects of the device could be improved with further work.

The largest uncertainty exists in the control of the AM distortion (detector slope) and pressure shift; and efforts should be concentrated in this area. It should be noted that it is possible to use different approaches for these problems. For example, one could modulate the source at two different frequencies and with different amplitude; this would allow one to independently servo the detector slope to zero thus eliminating AM distortion entirely. (Similar to the 3rd and 5th harmonic locks used to servo the oscillator and cavity). One is then left with the raw pressure shift which could be compensated in the output frequency by, for example, applying a calibrated correction voltage at the integrator input based on the observed signal amplitude.

ACKNOWLEDGEMENTS

The authors wish to thank the other members of the Frequency and Time Standards Section of NBS for their help and encouragement. Particular thanks are given to H. Hellwig for the original idea of reviving the ammonia standard for special application; H.E. Bell gave much help in the construction of the physical components. We thank F.L. Walls, S.R. Stein, R.M. Garvey, S. Jarvis, Jr., A.S. Risley and T.C. English (Efratom Inc.) for useful ideas and encouragement during the course of the work.

REFERENCES

- [1] WINELAND, D.J., HOWE, D.A. and HELLWIG, H., "Special Purpose Atomic (Molecular) Standard," to be published in the Proc. 8th Annual Precise Time and Time Interval (PTTI) Planning Meeting, December 1976.
- [2] LYONS, H., *Annals New York Acad. Sci.*, **55**, 831 (1952).
- [3] ALLAN, D.W., "Statistics of Atomic Frequency Standards," *Proc. IEEE*, **54**, February 1966, pp. 221-230.
- [4] HODOWANEC, G., "Microwave Transistor Oscillators," RCA Application Note AN 6291.
- [5] OXNER, E., "High Performance FETs in Low Noise Oscillators," Siliconix Incorporated, December 1973.
- [6] TOWNES, C.H., SCHAWLOW, A.L., *Microwave Spectroscopy*, McGraw-Hill Book Co., New York, 1955.
- [7] De PRINS, J., MENOUD, C., and KARTASCHOFF, P., " $N^{15}H_3$ Double Beam Maser as a Primary Frequency Standard," *Rev. of Sci. Instr.* Vol. 32, No. 11, Nov. 1961, pp. 1267-1268.
- [8] De PRINS, J., " $N^{15}H_3$ Double Beam Maser as a Primary Frequency Standard," *IRE Trans. on Instr.*, Vol. I-11, Dec. 1962, pp. 200-203.
- [9] KUKOLICH, Steven G., "Measurement of Ammonia Hyperfine Structure with a 2-Cavity Maser," *Phys. Rev.*, Vol. 156, No. 1, April 1967, pp. 83-92.
- [10] HAYWARD, D.O., and TRAPNELL, B.M.W., *Chemisorption*, 2nd Edition. Butterworths, Washington, 1964, pp. 205-248.
- [11] VIENNET, J., AUDOIN, C. and DESAINTEFUSCIEN, M., "Cavity Pulling in Passive Frequency Standards," *IEEE Trans. on Instrm. and Meas.*, Vol. IM-21, No. 3, August 1972, pp. 204-209.
- [12] The length of the waveguide between reflection points must be approximately equal to an integral number of half-wavelengths.
- [13] WALLS, F.L., "Design and Results from a Prototype Passive Hydrogen Maser Frequency Standard," to be published in the Proc. 8th Annual Precise Time and Time Interval (PTTI) Planning Meeting, Dec. 1976.
- [14] SHIMODA, K., "Atomic Clocks and Frequency Standard on Ammonia Line I," and "III" *J. Phys. Soc. Japan* **9**, June 1954, pp. 378-386 and pp. 567-575.
- [15] See for example: HELLWIG, H., "Atomic Frequency Standards: A Survey," *Proc. IEEE*, Vol. 63, No. 2, Feb. 1975, pp. 212-229. AUDOIN, C. and VANIER, J., "Atomic Frequency Standards and Clocks," *J. of Phys. E: Sci. Instr.* Vol. 9, 1976, pp. 697-720.
- [16] WALLARD, A.D., *J. Phys. E*, **5**, 927 (1972).
- [17] SHIRLEY, J.M., "Some Causes of resonant frequency shifts in atomic beam machines, II. The effect of slow frequency modulation on the Ramsey line shape," *J. Appl. Phys.*, Vol. 34, pp. 789-791, 1963.
- [18] A more exact treatment when both pressure and Doppler effects are important has been given by: PARSONS, R.W., and ROBERTS, J.A., "The Doppler

References (cont.)

Contribution to Microwave Line Widths," J. of Mole. Spec. 18, 1965, pp. 412-417.

- [19] HEWITT, P.L. and PARSONS, R.W., "Collision Broadening and Shifting in the Inversion Spectrum of NH_3 ," Phys. Lett., Vol. 45A, No. 1, Aug. 1973, pp. 21-22.

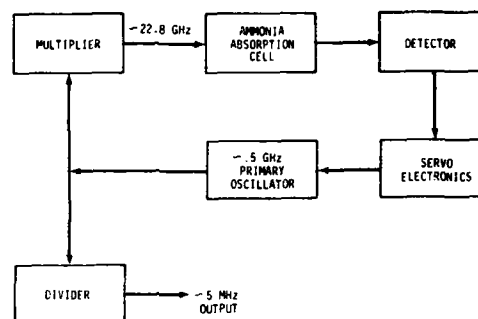


FIG. 1 Simplified block diagram of system. Frequency lock servo is used employing 0.5 - 25 kHz FM on ~ 5 GHz oscillator.

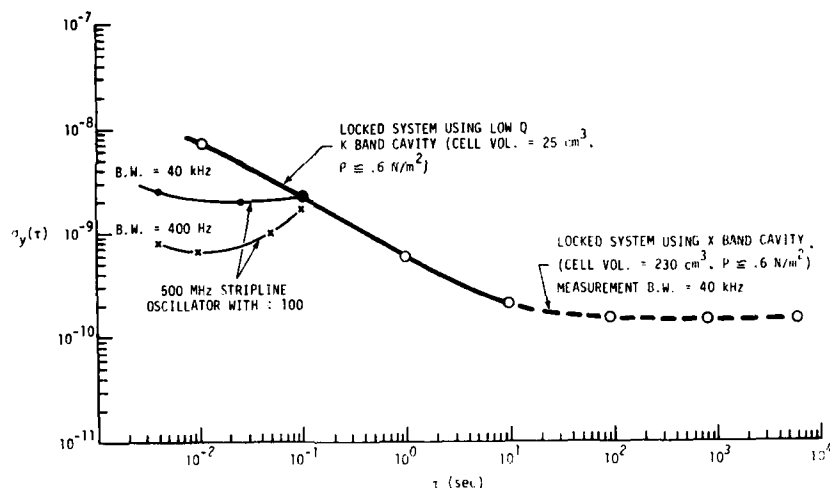


FIG. 2 Frequency stability plots showing the two sample variance (σ_y) for different averaging times (τ). B.W. \equiv measurement system bandwidth.

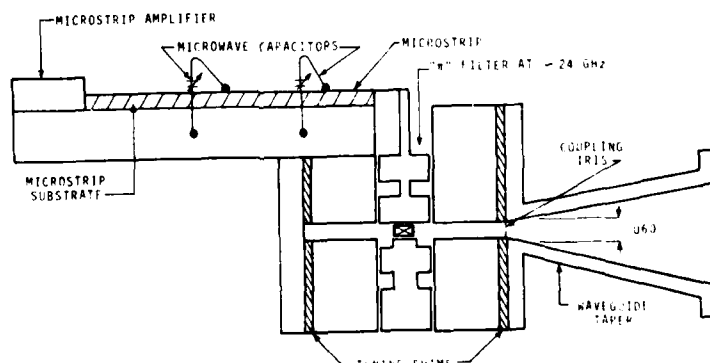


FIG. 3 Cross-section of step-recovery diode multiplier module.

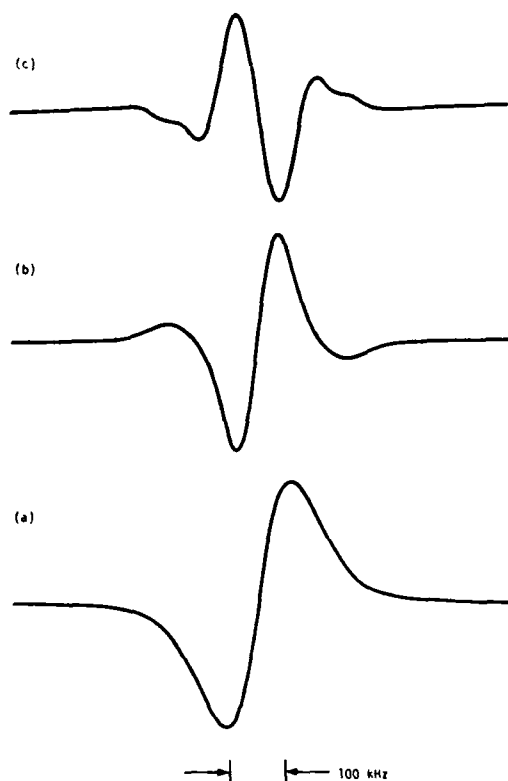


Fig. 4 Dispersion curves obtained at output demodulator for different harmonics: (a) Fundamental, (b) third harmonic, (c) fifth harmonic. Vertical scales are different in the three cases and curve (c) has an (arbitrary) 180 degree shift introduced.

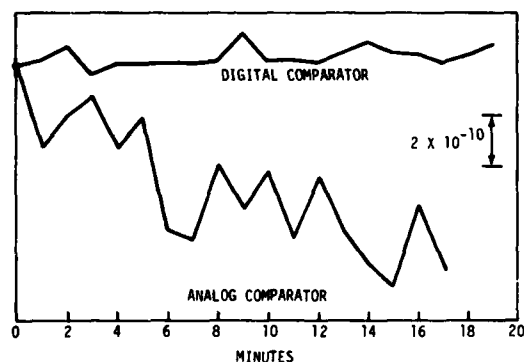


Fig. 7 Drift of analog and digital comparator (demodulator) in time. Vertical scale calibrated in terms of equivalent frequency offset (error) introduced by comparator.

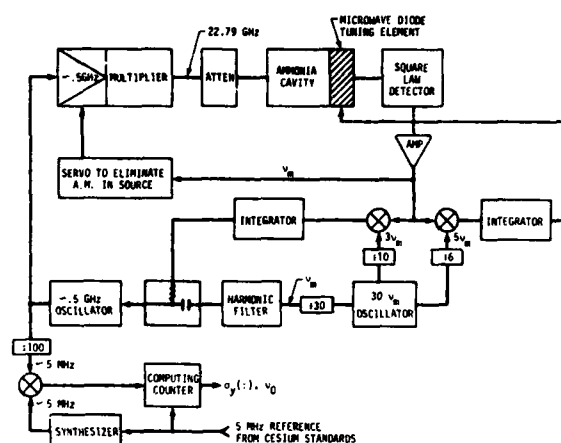


Fig. 5 Detailed block diagram of system.

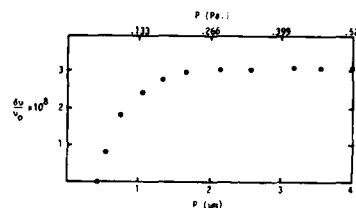


Fig. 6 Fractional frequency offset (from arbitrary reference) versus pressure. (Absolute pressure known to only a factor of 2.)

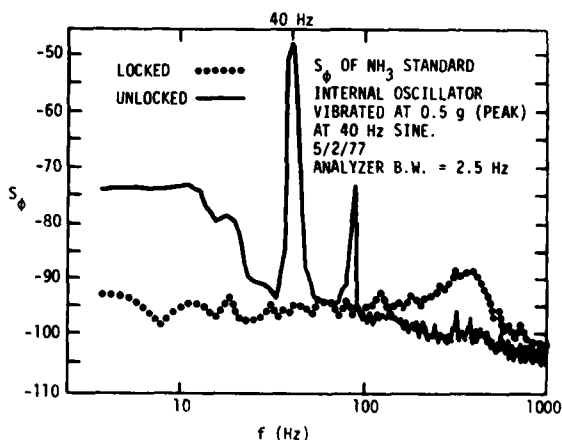


Fig. 8 Plot of phase spectral density $S_\phi(f)$ for locked and unlocked ~ 0.5 GHz oscillator.

DETECTING AND MIXING AT FIR/SUBMILLIMETER WAVELENGTHS WITH SUBMICRON SIZE SCHOTTKY BARRIER DIODES*

M. McColl, D. T. Hodges, A. B. Chase and W. A. Garber

The Ivan A. Getting Laboratories
The Aerospace Corporation
El Segundo, California 90245

Abstract

Progress on fundamental mixing experiments at submillimeter wavelengths using submicron size Schottky barrier diodes in an open mount is presented. Minimum detectable powers for the mixer of 3×10^{-14} W/Hz at 118.8 μ m and 1×10^{-17} W/Hz at 447 μ m are obtained. The result at 118.8 μ m represents a several order of magnitude improvement for a Schottky barrier mixer at this wavelength, and the 447 μ m result is the smallest to be reported for a Schottky diode employing an open mount. The rationale for using ultra small contacts is discussed.

Key words: Schottky Barrier Diode, Far Infrared Mixing, Submillimeter Wave Mixing, Electron Lithography.

Introduction

The Schottky barrier diode is the dominant mixer at microwave and millimeter wavelengths. It provides the widest bandwidths and lowest noise figures available from a room temperature device, and in addition, is both mechanically and electrically stable. For the past several years the Schottky diode has been under development for application at submillimeter wavelengths.¹⁻⁶ Most recently, small submicron dimensional Schottky diodes have been developed and extended to wavelengths as short as 70 μ m as a mixer and to a wavelength of 42 μ m as a video detector.^{3,4} This paper reports record sensitivities for these devices as mixers at wavelengths of 118.8 μ m and 447 μ m. The diodes were situated in an open untuned mount, the type of structure most commonly used in precision frequency synthesis and metrology experiments. Design considerations are discussed for diodes operating in such a mount.

The Schottky Barrier Diode

A Schottky-barrier diode consists of a metal contact to a semiconductor. The diode current voltage (I-V) characteristic is given to a good approximation by⁷

$$I = I_0 [\exp(SV) - 1] \quad (1)$$

where I_0 is determined by the area and material parameters of the diode. The parameter S is given by

$$S = \frac{q}{nkT} \quad (2)$$

where q is the electronic charge, T is the temperature, k is Boltzman's constant, and n is the ideality factor which is approximately unity for low- to moderately-doped materials. For heavily doped semiconductors where the conductance of the junction is dominated by tunneling, S is less than q/kT ; it is determined by materials parameters and doping, with only a weak dependence on temperature.^{8,9} This

highly nonlinear relationship between I and V expressed in (1) is responsible for the success of the Schottky diode as a detector and mixer.

The operation of the Schottky barrier detector at wavelengths considerably greater than the plasma resonance wavelength λ_p (approximately 5 μ m to 50 μ m for Si, Ge, and GaAs, depending on the material and doping)¹⁰ is best understood by examining the equivalent circuit of the diode shown in Fig. 1a.¹¹ The element R represents the nonlinear junction resistance that provides the rectifying volt-ampere behavior necessary for detecting and mixing. The spreading resistance R_s is the resistance in the bulk of the semiconductor which results from the constriction of the current near the metal contact. Z_{ss} is the skin effect resistance. The equivalent circuit shown in Fig. 1b becomes appropriate for wavelengths near λ_p where carrier inertia and displacement current effects within the semiconductor become dominant.¹² At λ_p , Z_s is resonant and approaches an open circuit condition.

The parasitic elements Z_s , Z_{ss} , and the junction capacitance C are the primary cause of the degradation in diode performance at high frequencies. All three elements are unavoidable but can be minimized by choosing the proper combination of materials, geometry, and especially doping. Heavily doped 5×10^{18} cm⁻³ n-type GaAs is a high mobility, high conductance semiconductor, and as such R_s and Z_{ss} are minimized by its usage.¹³ Moreover, raising the doping of a material extends the plasma resonance to shorter wavelengths. (For 5×10^{18} cm⁻³ n-GaAs, plasma resonance is at 13 μ m rather than 34 μ m for moderately doped n-GaAs.)¹⁰ Heavy doping also minimizes transit time effects.^{14,15} As a consequence of these considerations, 5×10^{18} cm⁻³ n-GaAs is the material utilized in the fabrication of diodes used in these experiments and those previously reported.^{3,4}

The relative importance of the parasitic elements depends upon the mount in which the diode is situated. In open mounts, C is not easily tuned out (resonated) and at short wavelengths becomes a dominant parasitic since it functions to bypass the high frequency RF currents from flowing through R. Consequently at submillimeter wavelengths, good performance from a Schottky diode with an open mount dictates that C should be small.

Reduction of the size of the diode improves its performance as a detector and mixer since junction capacitance is proportional to junction area. This is the basic approach taken in this work. There is a limit in the size of the diode in terms of impedance matching¹⁶ and electrical stability (burnout) problems; however, it is clear that diode diameters less than 1 μ m are preferred at submillimeter wavelengths.

Conversion Loss Dependence on Contact Diameter

The conversion loss L_c of a resistive mixer operated in an open mount under local oscillator starved conditions may be conveniently expressed as

*This work was supported by The-Aerospace Corporation on company-financed funds.

the product of three terms

$$L_c = L_0 L_1 L_2. \quad (3)$$

The intrinsic conversion loss L_0 is the loss arising from the conversion process within the nonlinear resistance of the diode.^{11,16} L_1 is defined here as representing the RF losses associated with scattering and reradiation of the signal power, RF impedance mismatch, RF parasitic losses, poor focusing, etc. The loss term L_2 contains the IF parasitic and IF impedance mismatch losses.

Local oscillator (LO) starvation refers to the mixer being driven with low LO power, a common situation at submillimeter wavelengths, particularly in open mounts. A convenient expression for L_0 in this condition is given by^{11,16}

$$L_0 = \frac{8}{S^2 R P_{LO}} \quad (4)$$

where

$$R = \frac{1}{S I_{dc}}. \quad (5)$$

R is the dynamic resistance of the diode biased with a dc current I_{dc} , and P_{LO} is the LO power absorbed by the mixer. Equation (4) is valid for $SV_1 \ll 1$ where $V_1 = \sqrt{2 R P_{LO}}$ is the amplitude of the LO sinusoidal voltage appearing across the resistance R .

Defining P_{LO} as the power in the incident LO beam and assuming the RF coupling problems are identical for both the signal and LO beams, one has

$$P_{LO} = \frac{P'_{LO}}{L_1}. \quad (6)$$

Hence from (3), (4) and (6)

$$L_c = \frac{8}{S^2 R P'_{LO}} L_1^2 L_2. \quad (7)$$

This equation leads to a striking relationship between L_c and diameter d when one examines the dependence of L_1 on d under LO starved conditions.

When the detector is capacitance-limited (i.e., $\omega^2 C^2 R^2 \gg 1$), the RF power fed to R is proportional to $(\omega C)^{-2}$. Hence, the dependence of L_1 on C becomes

$$L_1 \propto \omega^2 C^2; \quad \omega^2 C^2 R^2 \gg 1. \quad (8)$$

Expressing L_1 in terms of diameter d , one concludes $L_1 \propto d^4$. Preliminary responsivity measurements reported below lend some credence to this dependence. When a Schottky mixer is both LO starved and capacitance limited, Eqs. (7) and (8) lead to

$$L_c \propto \omega^4 C^4 \propto d^8; \quad \omega^2 C^2 R^2 \gg 1 \text{ and } SV_1 \ll 1. \quad (9)$$

These strong dependences of L_c on ωC and d have not, as yet, been verified by direct experiment. However, theoretically it is clear from Eq. (9) that reducing d should be an effective method of reducing L_c at short wavelengths.

Diode Fabrication

The diodes used in the measurements consisted of plated Pt contacts on uniformly-doped, non-epitaxial $5 \times 10^{18} \text{ cm}^{-3}$ n-type GaAs. Two different sizes were fabricated, 0.5 μm and 0.25 μm in diameter. A 0.5 μm Schottky contact with this doping has a calculated zero-bias capacitance C_0 of $1.3 \times 10^{-15} \text{ F}$. This value and a measured dc series resistance R_s of 19 ± 3 ohms yields a calculated figure of merit cut-off frequency $f_c = (2\pi R_s C_0)^{-1}$ of $9 \times 10^{12} \text{ Hz}$ (a wavelength of 34 μm). The 0.25 μm diodes were fabricated using high field pulse plating. This new method of plating has been demonstrated to yield spreading resistance values close to their theoretical values over a range of diameters from 0.12 μm to 2.5 μm .¹⁷ Incorporating this plating technique in the fabrication of the 0.25 μm diameter junctions has yielded a calculated f_c value of 40 THz (a wavelength of 7 μm).

The submicron size junctions were fabricated using electron lithographic techniques.¹⁸ Relatively large arrays of 1000 \AA diameter diodes of the type shown in Fig. 2 can be produced routinely.⁴ These diodes represent the smallest Schottky barrier diodes yet fabricated and possibly represent the smallest devices, in general, to be reported in the literature.

Conversion Loss and Noise Measurements

The sensitivity of Schottky diodes in open-untuned mounts was measured using signal and local oscillator sources derived from separate lasers. The radiation was coupled to the diode through a whisker (long wire) antenna which also functioned as the electrical contact to the diode. Optimum signal was obtained using a whisker orientation which selectively excites the main lobe of the long wire antenna.¹⁹ The signal and local oscillator consisted of two optically pumped far infrared waveguide lasers.^{20,21} The beams were combined using a beam splitter consisting of a wire grid polarizer with grid lines oriented at approximately 45° to the laser polarization. With both lasers oscillating on the same transition, intermediate frequencies up to 2 MHz at 447 μm and 5 MHz at 118.8 μm could be generated by tuning each far infrared laser cavity off resonance. The detected signal was fed to a 10 MHz IF amplifier (26 dB gain) and displayed using an HP 8553 spectrum analyzer.

The power available at the signal and local oscillator inputs was measured with a Scientech Model 360 calorimeter. This is a bulk absorbing device which has been calibrated to an accuracy of $\pm 20\%$ for the wavelengths used in these measurements.^{20,21} Following combination in the beam splitter, the signal and local oscillator beams were focused onto the whisker with a 5 cm focal length spherical mirror. Coupling to the whisker was degraded by the present geometrical constraint of using the spherical mirror off axis. The shadowing due to the diode mount precludes an on-axis arrangement. In addition, the spatial nonuniformity (higher order modes) of both signal and local oscillator prevented good beam overlap at the focus.

The results of conversion loss measurements at wavelengths of 118.8 μm and 447 μm using the 1/2 μm diameter diodes at a dc bias current $I_{dc} = 100 \mu\text{A}$ are shown in Table I. The L_c values represent the ratio (in dB) of the signal laser beam power to the power fed to an impedance matched IF amplifier. These data were obtained under conditions in which the mixer was sufficiently LO starved that both the signal impedance and the IF impedance of the device was approximately independent of LO drive, and hence,

from (5) both were equal to $(S I_{dc})^{-1} = 400$ ohms, where $S = 25 \text{ V}^{-1}$. As such, the input impedance of the amplifier was adjusted to equal 400 ohms. The LO laser beam power P_{LO} employed is shown in the second column.

The minimum detectable signal power MDP_M of a mixer referred to its input is given by^{5,22}

$$MDP_M = k L_c T_D B_{IF} \quad (10)$$

where T_D is the noise temperature of the mixer diode measured at the IF port and B_{IF} is the IF bandwidth. Measurements of T_D were taken at frequencies of 1, 10, 30 and 105 MHz with $I_{dc} = 100 \mu\text{A}$. A strong $1/f^m$ noise spectrum with $m \approx 1$ is observed at the three lower frequencies that would indicate a noise corner frequency f_n of 50 MHz with $I_{dc} = 100 \mu\text{A}$. (Noise corners of a few tens of megahertz are not uncommon with Schottky diode detectors.)²³ At 105 MHz, a T_D of approximately 300°K is obtained. T_D is found independent of LO power at these low levels of LO drive; this is consistent with previous observations.^{24,25} From these noise and conversion loss measurements, the MDP_M values shown in Table I are calculated using (10) with $T_D = 300^\circ\text{K}$. As such, these values apply only for IF frequencies above 50 MHz.

The measurements reported in this paper were confined to the $1/2 \mu\text{m}$ diameter diodes. Preliminary video responsivity measurements have also been performed on the $1/4 \mu\text{m}$ diameter diodes. A comparison of these values with those obtained for the $1/2 \mu\text{m}$ diameter diodes under similar conditions of mount, whisker antenna, etc., does indicate improved coupling efficiencies with the $1/4 \mu\text{m}$ diode. However, the improvement is less than the factor of 16 predicted by (8) and could be due to errors in measuring diode size and fringing effects from the whisker point. Work is now in progress to clarify the operating characteristics of the $1/4 \mu\text{m}$ diodes.

Acknowledgments

The authors wish to thank R. D. Reel for his aid in the conversion loss measurements and R. E. Robertson and W. DeLaTorre for their help in the fabrication of the diodes. Appreciation is due W. A. Johnson for supplying the low noise amplifiers and for helpful discussions on noise measurement.

References

1. H. R. Fetterman, B. J. Clifton, P. E. Tannenwald, and C. D. Parker, "Submillimeter detection and mixing using Schottky diodes," *Appl. Phys. Lett.*, vol. 24, pp. 70-72, Jan. 1974.
2. H. R. Fetterman, B. J. Clifton, P. E. Tannenwald, C. D. Parker, and H. Penfield, "Submillimeter heterodyne detection and harmonic mixing using Schottky diodes," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-22, pp. 1013-1015, Dec. 1974.
3. D. T. Hodges and M. McColl, "Extension of the Schottky barrier detector to $70 \mu\text{m}$ (4.3 THz) using submicron dimensional contacts," *Appl. Phys. Lett.*, vol. 30, pp. 5-7, Jan. 1977.
4. M. McColl, D. T. Hodges and W. A. Garber, "Submillimeter-wave detection with submicron size Schottky barrier diodes" *IEEE Trans. Microwave Theory Tech.*, vol. MTT-25, pp. 463-467, June, 1977.
5. M. McColl, "Review of Submillimeter wave mixers", *Proc. of the SPIE: Far Infrared/Submillimeter Wave*, vol. 105, 1977 (in press).
6. H. R. Fetterman, "Advanced Schottky diode concepts," *Proc. of the SPIE: Far Infrared/Submillimeter Wave*, vol. 105, 1977 (in press).
7. S. M. Sze, *Physics of Semiconductor Devices*, New York: Wiley, 1969, pp. 393-394.
8. F. A. Padavani and R. Stratton, "Field and thermionic-field emission in Schottky barriers," *Solid-State Electron.*, vol. 9, pp. 695-707, 1966.
9. M. F. Millea, M. McColl and C. A. Mead, "Schottky barriers on GaAs," *Phys. Rev.*, vol. 177, pp. 1164-1172, Jan. 1969.
10. For a survey on plasma resonance, see P. A. Schumann, "Plasma resonance calibration curves for silicon, germanium and gallium arsenide," *Solid State Tech.*, vol. 13, pp. 50-51, Jan. 1970.
11. H. C. Torrey and C. A. Whitmer, *Crystal Rectifiers* (M.I.T. Radiation Lab. Ser., vol. 15), New York: McGraw-Hill, 1948.
12. K. S. Champlin and G. Eisenstein, "Cut-off frequency of submillimeter Schottky-barrier diodes," submitted for publication.
13. M. McColl, M. F. Millea, J. Munushian, and D. F. Kyser, "Improved 94-GHz GaAs mixer diodes using gold-copper alloy whiskers," *Proc. IEEE*, vol. 55, pp. 2169-2170, Dec. 1967.
14. K. K. Thornber, T. C. McGill and C. A. Mead, "The tunneling time of an electron," *J. Appl. Phys.*, vol. 38, pp. 2384-2385, April 1967.
15. A. van der Ziel, "Infrared detection and mixing in heavily doped Schottky-barrier diodes," *J. Appl. Phys.*, vol. 47, pp. 2059-2068, May 1976.
16. M. McColl, "Conversion loss limitations on Schottky-barrier mixers," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-25, pp. 54-59, Jan. 1977.
17. M. McColl, A. B. Chase and W. A. Garber, "Electrical behavior of submicron GaAs Schottky barrier diodes," *Bull. Am. Phys. Soc.*, vol. 22, pp. 260-261, March 1977.
18. M. McColl, W. A. Garber, and M. F. Millea, "Electron beam fabrication of submicrometer diameter mixer diodes for millimeter and submillimeter wavelengths," *Proc. IEEE*, vol. 60, pp. 1446-1447, Nov. 1972.
19. L. M. Matarrese and K. M. Evenson, "Improved coupling to infrared whisker diodes by use of antenna theory," *Appl. Phys. Lett.*, vol. 17, pp. 8-10, July 1970.
20. D. T. Hodges, F. B. Foote, and R. D. Reel, "Efficient high-power operation of the CW far-infrared waveguide laser," *Appl. Phys. Lett.*, vol. 29, pp. 662-664, Nov. 1976.

21. D. T. Hodges, F. B. Foote, and R. D. Reel, "High Power Operation and Scaling Behavior of CW Optically Pumped FIR Waveguide Lasers," IEEE J. Quant. Electr., June 1977 (in press).
22. M. McColl, R. J. Pedersen, M. F. Bottjer, M. F. Millea, A. H. Silver and F. L. Vernon, Jr., "The super-Schottky diode microwave mixer," Appl. Phys. Lett., vol. 28, pp. 159-162, Feb. 1976.
23. H. E. Elder and V. J. Glinski, "Detector and mixer diodes and circuits," in Microwave Semiconductor Devices and Their Circuit Applications, edited by H. A. Watson, New York: McGraw-Hill, pp. 370-395, 1969.
24. D. T. Young and J. C. Irvin, "Millimeter frequency conversion using Au-n-type GaAs Schottky barrier epitaxial diodes with a novel contacting technique", Proc. IEEE, vol. 53, pp. 2130-2131, Dec. 1965.
25. R. J. Bauer, M. Cohn, J. M. Cotton, Jr., and R. F. Packard, "Millimeter wave semiconductor diode detectors, mixers, and frequency multipliers," Proc. IEEE, vol. 54, pp. 595-605, April 1966.

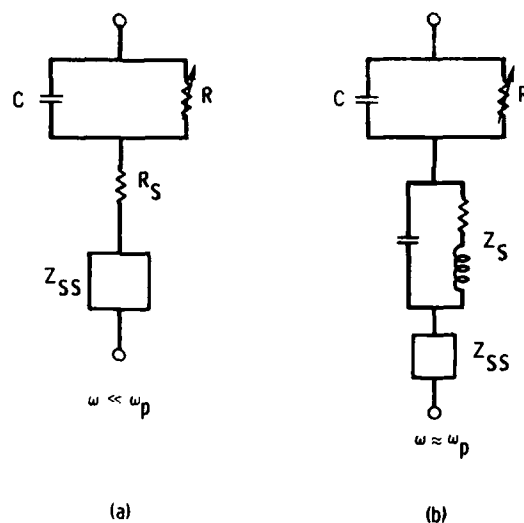


FIGURE 1. The equivalent circuits of a Schottky barrier diode for (a) frequencies considerably less than the plasma resonance frequency and (b) frequencies near the plasma frequency of the semiconductor.

λ (μm)	P_{LO}^1 (mW)	L_C (dB)	MDP_M (W/Hz)
118.8	13	69	3×10^{-14}
447	4	34	1×10^{-17}

TABLE 1. Mixing results using $\frac{1}{2}$ diameter Schottky barrier diodes.

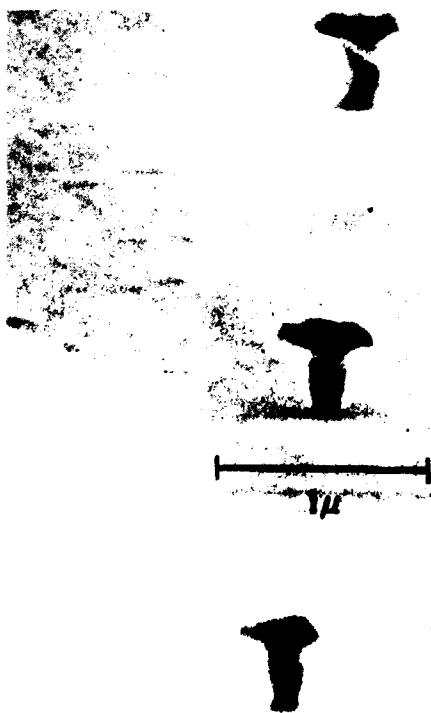


Fig. 2a



Fig. 2b

Fig. 2 Scanning electron microscope photographs of plated Schottky diodes. The insulating layer normally present has been stripped away. The photographs were taken at steep angles from normal incidence to show the dimensions of the diodes at their base. Figure 2b is a magnified shot of one of the diodes shown in Fig. 2a.⁴

A NARROW OUTPUT LINEWIDTH MULTIPLIER CHAIN
FOR PRECISION FREQUENCY MEASUREMENTS IN THE 1 THz REGION

E. Bava, A. De Marchi, A. Godone
Istituto Elettrotecnico Nazionale "Galileo Ferraris"
Torino, Italy

Summary

A multiplier chain is described which is a substantially improved version of the one previously used at IEN. Starting at 5 MHz, a single stage x12 multiplier which makes use of a low frequency step recovery diode provides at 60 MHz a 1 mW signal suitable for narrow band filtering in a 10 kHz wide quartz filter. In this way, from oscillators with white phase noise floor of -140 dB, a coherent "carrier" signal is obtained at half a THz with a signal-to-noise ratio of about 10 dB (in 1 kHz bandwidth) over a 50 kHz wide pedestal.

The use of 5 MHz oscillator with better phase stability makes our chain become the limiting factor, with a white phase noise floor of -165 dB, and pushes the carrier collapse frequency of the system up to 10 THz. A signal-to-noise ratio of 35 dB (in 1 kHz bandwidth) is then obtained at half a THz with a 25 kHz wide pedestal and no observable carrier linewidth with the selectivity used.

Further work on the multiplier may bring the carrier collapse frequency of the system with these new oscillators as high as 30 THz or more. The power obtainable from this chain at the different harmonics of the used 100 GHz klystron is not impressive (as low as 100 nW at 0.6 THz) but the instrument is suitable for precise absolute frequency measurements and high resolution spectral analysis in the THz region.

Introduction

The generation by synthesis of signals in the far and near infrared with a usable power and high stability and spectral purity is a problem which is being investigated in many laboratories around the world.^{1,2} The reasons for this are both scientific and applicative.

In fact the availability, in that region of the spectrum, of very narrow linewidth signals, whose frequency is directly connected with the existing frequency standards (Cs beam tube, H maser):

- a) would make possible the stabilization of sources in the infrared providing a chain

of coherent equal stability standards for a metrological frequency scale;

- b) would allow improvements by several orders of magnitude in the precision of frequency measurements of spectral lines, giving a suitable absolute reference to the modern field of very high resolution sub-Doppler spectroscopy;
- c) could indicate a spectral line of some substance with superior stability characteristics with respect to the Cs resonance and eventually bring to a new frequency standard with better performances in a region where it could be also a wavelength standard.

Moreover the extension to higher frequencies of the region where precision absolute frequency is possible would offer to the communication technology new possibilities for very wideband channels.

The problems that arise in the generation of high frequency signals are essentially two. One is that the power obtainable from harmonic distortion in non-linear devices is low and high number harmonics are usually difficult to generate (except for the Josephson junction); this difficulty however can be overcome, where a tunable source is available, by (offset) locking it to the multiplied signal.

The other is that the spectral purity of the synthesized signal depends on the phase noise power spectrum of the source used so that the signal-to-noise ratio S/N decreases with the square of the multiplication number.

In addition, when the multiplication number is such that the power in the noise sidebands becomes comparable to the signal power, the power in the carrier P_c decreases dramatically and the noise pedestal itself begins to widen.^{3,4}

For any given synthesizer a characteristic carrier collapse frequency ν^{col} may be introduced for which the carrier power is reduced of a factor e^{-1} . In order to increase the value of ν^{col} the total phase noise power Φ^P at a given frequency must be decreased, and this can be done only by reducing the bandwidth of

the system or by reducing the noise of the source (or of the multiplier, whichever is bigger).

The two things are obviously not equivalent since the first one just increases the carrier collapse frequency but does not affect S/N for $\sqrt{f} < \sqrt{f_{col}}$, while the second increases both $\sqrt{f_{col}}$ and the S/N ratio.

Both ways have been experienced at IEN in the attempt of improving, with respect to the results previously obtained, the spectral purity of the synthesized signals in the region below 1 THz.⁴ The performances achieved with the new synthesizers are described in this paper.

VHF section of the chains

Since the phase noise of a multiplied signal increases along the chain, the most important contributions to it come from the source and from the first stages of the multiplier.

For this reason we selected the oscillators with the lowest phase noise available on the market, and spent some efforts in improving the noise of our chains.

The spectral distributions of phase noise power in the best oscillators available at IEN are shown in fig. 1 (curves a and b) together with the phase noise, reported to the input, of two 5 MHz - 60 MHz chains (curves c_1 and c_2).

When the oscillators corresponding to curve a were used no disturbance was brought by the phase noise of the multiplier, so that the only way to increase $\sqrt{f_{col}}$ was to reduce the bandwidth of the system, which we did by inserting a quartz crystal filter with a -3 dB half bandwidth of 12 kHz at 60 MHz, just after the first stage of the chain, as it is shown in the block diagram of fig. 2; fig. 3 gives the transfer function of the filter.

This arrangement led to a carrier collapse frequency of 500 GHz, which is a factor of 3 higher than the one obtained previously.²

However, when new 5 MHz oscillators with better spectral purity became available (curve b) with a potential $\sqrt{f_{col}}$ of 50 THz if quartz filtered, we made some trials to improve the noise of our multipliers in order to exploit them better.

The first x 12 step was realized both with a single push-pull bipolar transistor stage, and with a step recovery diode (SRD) used in the circuit of fig. 4.

The phase instabilities of these two circuits are given by curves c_1 (the cut-off of

the crystal filter is not shown here) and c_2 in fig. 1; the SRD shows an improvement in the white noise floor, probably due to the simplicity of the circuit and the absence of any other resistive component than the diode, but an increase of the flicker noise, probably due to the recombination of carriers in the SRD.

With the new oscillators and the quartz filtered SRD stage a carrier collapse frequency of 10 THz was obtained. Work is in progress on a varactor stage from which superior characteristics are expected both in the phase noise and in the AM-PM conversion.⁵

All the three schemes provide indeed at their outputs a power level suitable for a quartz filter.

After crystal filtering and amplification at 60 MHz the power necessary for driving a SRD to reach X band is provided at 480 MHz by three cascaded doublers: one push-pull bipolar transistor stage and two overlay transistor multipliers.

Spectral analysis of the synthesized signals

A few experiments were carried out, both with the old and the new oscillators, by analyzing with coherent sources the signal produced by the chain at different synthesized frequencies.

The pedestal linewidth $\Delta\nu_p$ and the signal-to-noise ratio S/N can be in this way compared with the values predicted by the theory for the measured data of the chains and oscillators used.

Fig. 5 shows the beat note obtained from a Schottky barrier diode by mixing the signal from the chain with the output of an optically pumped CH_3F laser at 604 GHz built at the University of Pisa in a cooperation research aiming at the experimental analysis of the feasibility of a submillimetric atomic frequency standard.

The oscillator and multiplier phase noise spectra correspond in fig. 1 to curves a and c_2 respectively.

The IF bandwidth was 300 Hz in this registration and a small coherent signal could be seen in a linear display over a pedestal approximately 50 kHz wide. $\Delta\nu_p$ was evaluated by comparison with other logarithmic registrations. The insertion of the quartz filter at 60 MHz allowed in this way a significant improvement of the synthesized spectrum at 600 GHz (see previous registrations⁴).

When the new quartz oscillator was used, the phase noise of the chain became the basic

limitation for the spectral purity in the far infrared region.

Fig. 6 shows the spectrum of this new synthesizer observed at 9.6 GHz with a 1 kHz IF bandwidth.

From this chain a 300 MHz beat note was obtained in a Josephson point contact diode by beating the 97th harmonic of the 9.18 GHz signal from the chain with the 890.760 GHz output of an HCN laser.

This measurement was made at the "Institut d'Electronique Fondamentale" in Orsay, France, using an HCN laser and the Josephson mixer, developed in that laboratory.

A couple of mW from the laser and 4 mW from the chain were fed into the mixer.

Unfortunately the low power level of the beat note prevented us from achieving a complete determination of S/N and $\Delta\nu_p$. S/N was estimated to be more than 12 dB with 10 kHz IF bandwidth.

Fig. 7 shows the spectrum of a beat note obtained between the chain and an optically pumped (R (18) line of a CO₂ laser) formic acid laser built at PTB in Braunschweig, Germany. In this occasion a Schottky barrier mixer was used in an open structure.

The frequency of the laser line was measured to be 761.6077 GHz and S/N resulted to be about 22 dB with a 3 kHz IF bandwidth; $\Delta\nu_p$ was about 30 kHz.

Fig. 8 is a summary of the theoretical curves and the experimental results obtained for the chains and oscillators available at IEN. The ratio between the carrier power and the pedestal white phase noise level in a bandwidth of 100 Hz (sloping lines), and the pedestal -3 dB linewidth $\Delta\nu_p$ (sub-horizontal upper lines) are plotted as functions of the synthesized frequencies. The curves are derived from the theoretical formulas derived from the hypothesis that S_D and S_{RF} are equally shaped and asymptotically identical for high Fourier frequencies⁴:

$$\Delta\nu_p = 2B \left(\frac{\Phi'}{1 - e^{-\Phi'}} \right)^{\frac{1}{2-1}}$$

$$\left(\frac{S}{N} \right)_{BW} = 2 \frac{1 \ln \frac{\pi}{2}}{\frac{\pi}{2}} \frac{B}{BW} \Phi' \frac{1}{2-1} \frac{e^{-\Phi'}}{(1 - e^{-\Phi'})^{\frac{1}{2-1}}}$$

The curve b' refers to the best quartz using an ideal multiplier filtered with a 25 kHz wide crystal filter which rolls off as

f⁻⁴; the other symbols used in fig. 8 are explained in the caption.

The circles are experimental points and the arrows represent experimentally determined limitations.

Conclusions

A few experimental results have been presented on the spectral purity obtained in synthesized frequencies in the far infrared region. Our multiplier chains must be improved in order to exploit completely the high spectral purity of new 5 MHz crystal quartz oscillators which can assure in principle multiplied coherent signals beyond 30 THz (CO₂ laser). Beside the direct frequency measurement of sources in the far infrared, the phase lock of molecular lasers as well is allowed by means of these chains.

Acknowledgments

We want to express our appreciation to P. Minguzzi and his coworkers of Pisa University, to C.O. Weiss and G. Kramer of PTB Braunschweig, to R. Adde and J.J. Jimenez of the "Laboratoire Primaire du Temps et des Fréquences", for their kind cooperation that allowed us to obtain an experimental verification of the spectral purity theory for multiplied signals.

References

- 1 D.A. Jennings, F.R. Petersen and K.M. Evenson: Extension of absolute frequency measurements to 148 THz: Frequencies of the 2.0 - and 3.5 μ m Xe laser. Appl. Phys. Lett. 26, pp. 510 - 511 (1975).
- 2 T.G. Blaney, N.R. Cross and D.J.E. Knight: Harmonic mixing and frequency Measurement at 2.5 THz using Josephson Junctions. J. Phys. D: Appl. Phys. 9, pp. 2175 - 80 (1976).
- 3 F.L. Walls and A. De Marchi: RF Spectrum of a Signal after Frequency Multiplication: Measurement and Comparison with a Simple Calculation. IEEE Trans. on IM - 24 no. 3, pp. 210 - 217 (September 1975).
- 4 E. Bava, A. De Marchi and A. Godone: Spectral Analysis of Synthesized Signals in the mm Wavelength Region. IEEE Trans. on IM - 26 no. 2 (June 1977).
- 5 E. Bava, G.P. Bava, A. De Marchi and A. Godone: Measurement of Static AM - PM Conversion in Frequency Multipliers. IEEE Trans. on IM - 26 no. 1 (March 1977).

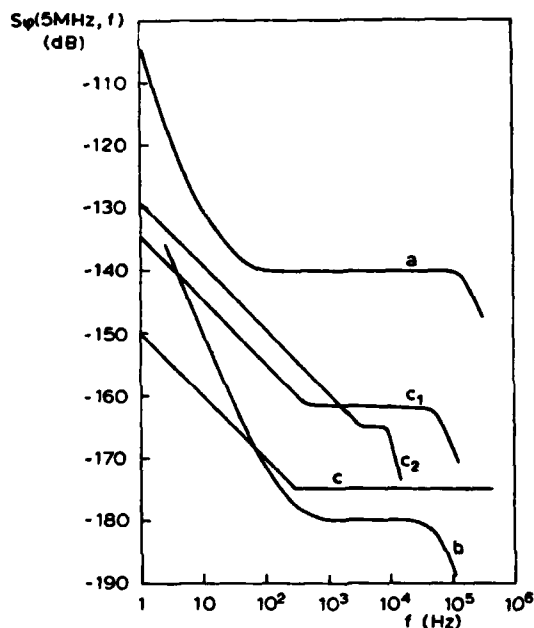


Fig. 1 - Spectral densities in rad^2/Hz of phase fluctuations:
a), b) 5 MHz crystal controlled oscillators;
c) phase noise of IEN measurement circuit;
 c_1 , c_2) noise of the chains measured at 60 MHz and referred to the input, respectively for transistor and step-recovery diode multiplier.

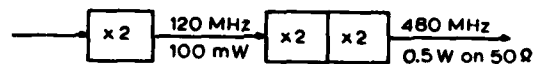
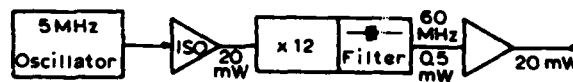


Fig. 2 - Block diagram of the multiplier chain from 5 MHz to 480 MHz.

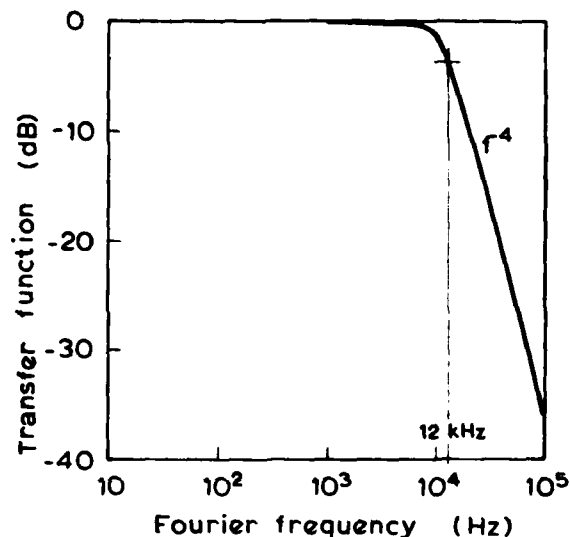


Fig. 3 - Transfer function of the 60 MHz quartz band-pass filter.

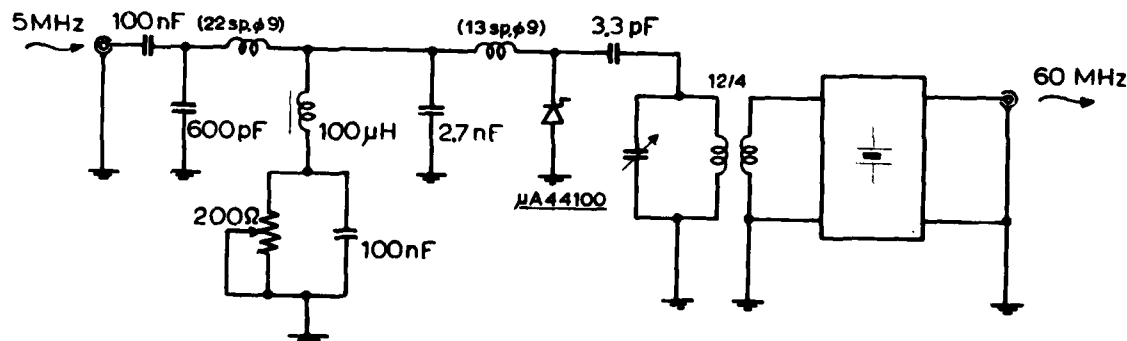


Fig. 4 - Electric circuit of the first multiplier stage with a step recovery diode and a quartz filter at the output.

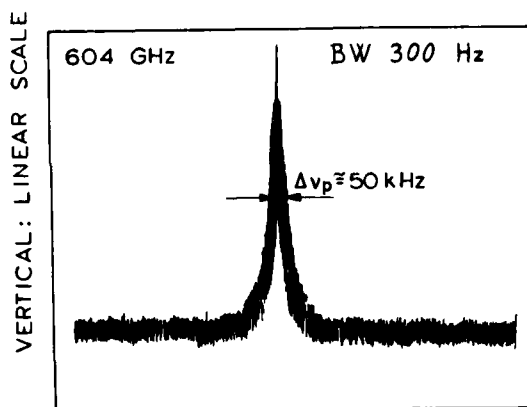


Fig. 5 - Spectrum of the beat note between a CH_3F laser (604.3 GHz) and the sixth harmonic of a millimetric klystron (100.3 GHz). S_ϕ of the 5 MHz quartz oscillator and multiplier chain shown in fig. 1, a and c_2 respectively. The IF bandwidth is 300 Hz.

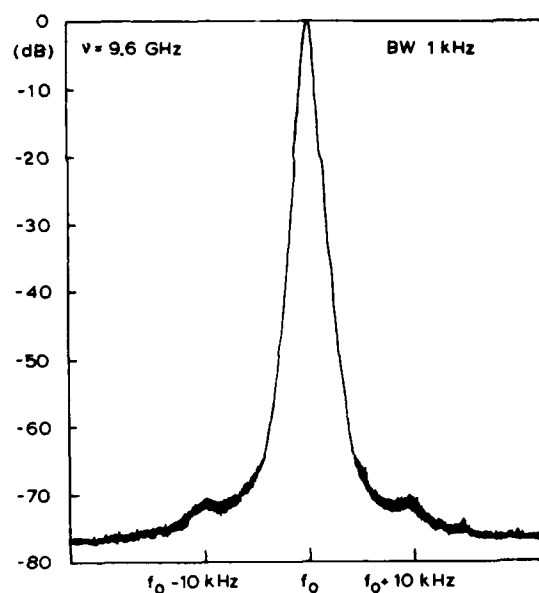


Fig. 6 - 9.6 GHz spectrum (1 kHz IF bandwidth) of the multiplier chain c_2 using an oscillator with S_ϕ given in fig. 1, b.

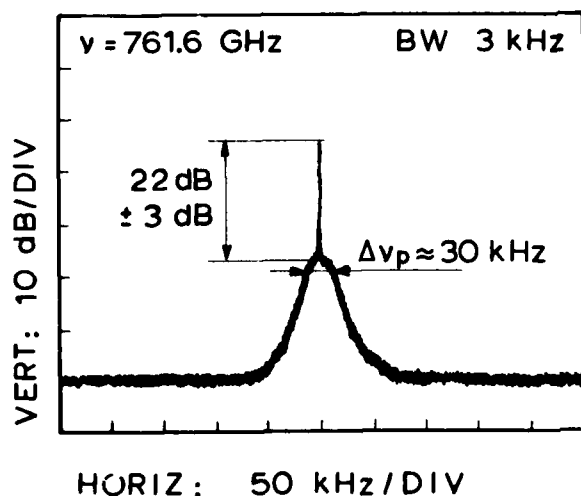


Fig. 7 - Spectrum of the beat note between a HCOOH laser (761.6 GHz) and the 11th harmonic of a millimetric klystron (69 GHz). The IF bandwidth is 3 kHz. Curves b and c_2 in fig. 1 give the S_ϕ spectrum of oscillator and chain respectively.

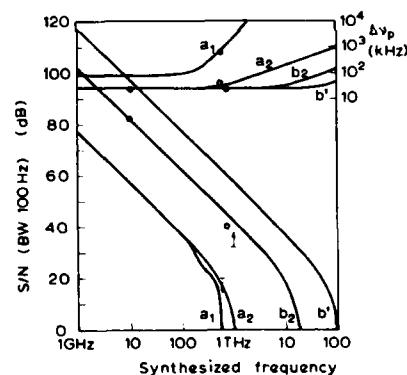


Fig. 8 - Curves of S/N in 100 Hz bandwidth and Δv_p for multiplier chains built at IEN a_1 , a_2 , b_2 and foreseen characteristic of the next chain b' . Circles refer to measured values, arrows refer to measurements with external factors limiting the evaluation of experimental data. With reference to S_ϕ of fig. 1.
 a_1 : combination of a and c_1 ;
 a_2 : combination of a and c_2 ;
 b_2 : combination of b and c_2 ;
 b' : combination of b and a chain with a filter having the same bandwidth as in c_2 and with a noise level not impairing that of source b.

A STUDY OF POINT-CONTACT JOSEPHSON JUNCTIONS FOR USE IN FREQUENCY SYNTHESIS

Allan S. Risley

Frequency & Time Standards Section
National Bureau of Standards
Boulder, CO 80302

ABSTRACT

A carefully controlled microwave experiment using point-contact Josephson junctions (PCJJ) is reported. The experiment measures the IF power, P_{IF} , generated in low-order frequency multiplication and mixing. Also measured is the number of induced steps produced by the microwave source. Four characteristics possessed by PCJJ's which gave the largest P_{IF} are described. These criteria are compared to those which were obtained in studies previously reported. There is good agreement with three of these criteria and the discrepancy with the fourth is discussed in detail. An operational procedure for fabricating a good PCJJ is described. The four most important potential high frequency limitations of a PCJJ are discussed with respect to the experimental results. This paper reports the use of a non-superconducting whisker plated with a superconductor.

INTRODUCTION

At the National Bureau of Standards we intend to phase-lock a far-infrared (FIR) laser to an X-band signal [1-3]. The aim is to be able to compare the frequency of a microwave source to that of a laser source, without losing cycle identification, at laser frequencies at least as high as 4.25 THz (70.5 μ m). The large frequency multiplication orders, N , required indicate that the point-contact Josephson junction (PCJJ) is presently the best candidate for the multiplier and mixer. This paper discusses a microwave experiment aimed at improving the understanding and reliability of PCJJ's.

While the PCJJ has performed $N = 401$ (starting at X-band) [4] and $N = 825$ (starting at 1 GHz) [5] the reliability is inadequate for the present application. The practical problem is to learn how to make PCJJ's of sufficient quality and reliability. Junctions capable of giving a signal-to-noise ratio of at least 20 dB (in a 1 kHz bandwidth) for $N = 500$ is the goal.

The empirical results of NBS [6] and those of NPL [7] evolved four criteria which appear to characterize a good PCJJ for use at large N and specifically, where the upper frequency source is in the FIR. They are:

- (1) The critical current, I_0 , should be of the order of 250 μ A. Critical currents lower than 50 μ A and higher than 1 mA are outside the optimum range.
- (2) The product, $I_0 R$, (where R is the asymptotic resistance of the junction) should be of the order of 1 mV or greater.
- (3) The undriven I-V curve should show evidence of an "energy-gap" structure. [The dc component of the total current through the junction is I and that component of the voltage across it is V .]
- (4) The junction should show at least the first step in the I-V curve due to the laser drive, V_{laser} .

Presumably these empirical observations have an important relationship to the four major features that potentially limit the N value and maximum value of V_{laser} . These limitations are:

- (a) The spreading of the spectrum of the supercurrent as N increases [8].
- (b) High-frequency limitations of the Josephson effect due to the finite binding energy (i.e., the energy gap) of superconducting electron pairs [9].
- (c) Localized heating (due to the driving sources) which lowers I_0 [10].
- (d) The shunting of high-frequency currents due to the non-zero capacitance of the contact [11].

II. THE EXPERIMENT

In the experiment reported here, both simple mixing ($N = 1$), and third harmonic generation and mixing ($N = 3$) are studied. Two X-band sources are used in the first case, and an X-band and a K-band source are used in the second. The measured quantities are the intermediate frequency power, P_{IF} , and the voltage extent of the steps in the I-V curve produced by the microwave sources [7, 12, 13].

Experience has shown - but not satisfactorily explained - the importance of several parameters to the generation of a useful level of P_{IF} . These are: the maximum supercurrent, I_0 ; the asymptotic resistance, R ; the product, $I_0 R$; and the dc component of the voltage across the junction, V [4, 14-17]. Figure 1 defines these quantities.

A block diagram of the experimental set-up is given in figure 2. A few comments need to be made about some of the components: When the microwave switch is in positions 2 or 3, the junction is irradiated with both microwave sources and the system is set up for measuring P_{IF} . When the switch is in position 1, the junction is only irradiated by source #1 and the system is set up for measuring steps in the I-V curve.

The break between the waveguide and coaxial tees, and the 1.4 GHz coaxial filter, provide isolation such that the PCJJ is the only dc path that the bias box sees. This allows a known bias to be put on the junction and permits observation of the I-V curve. In addition to this function, the coaxial filter provides impedance matching at ν_{IF} (= 1.4 GHz) and rejection of the two microwave drives.

The Josephson junction mount fits directly into an ordinary storage dewar and is coaxial in structure. The whisker for the junction forms the extreme end of the center conductor, and the post forms a short at the end of the shield. The probe is inserted into the dewar with the junction open (i.e., the whisker not contacting the post) and the junction is closed only after the device comes to temperature. This allows the formation of contacts of sufficient sensitivity to be of interest.

III. EXPERIMENTAL RESULTS

Table I displays the two variables, P_{IF} and voltage-extent of steps, versus the four parameters. Rather than list the total number of observable microwave steps, I have listed that value of V at which the last observable step occurs. [Steps whose full height was as small as about 5 μA were observable. With reduction of the noise fed to the junction it is possible to study steps as small as a few nanoamperes [18]. In retrospect, it is unfortunate that it was not possible to resolve smaller steps because this might have permitted a more definitive interpretation of the data (see Section IV)]. The reason for this is that the self oscillation frequency is directly proportional to V

($\nu_J = 484 \text{ MHz}/\mu V$) and the appearance of a step is evidence of a zero-beat between ν_J and a harmonic of the drive signal. The original thought was that the extent of the steps in any given junction would be a measure of the high-frequency cut-off of that junction. This does not, however, seem to be the case.

There are several main features to be noted in this table:

- (1) If P_{IF} is greater than about 40 dB above the noise (in a 30 kHz bandwidth), steps are not observed. If P_{IF} is less than about 40 dB there are (usually) well defined steps.
- (2) Based on 21 contacts, P_{IF} greater than 50 dB above the noise correlates fairly well with $I_0 R > 0.8 \text{ mV}$.
- (3) Based on 21 contacts, the absence of steps correlates rather well with $R > 2 \text{ ohms}$.
- (4) Based on 11 contacts, the voltage extent of the steps is not well correlated with I_0 .

Several other things are important to mention: At the time these data were taken, it was not possible to measure asymptotic resistance smaller than about 1 ohm. Thus, for many contacts, an accurate value of $I_0 R$ is not available and the corresponding $I_0 R$ element has been left blank.

An entry of N.M. in the P_{IF} column just means that P_{IF} was sufficiently low as not to be of interest, say, less than 35 dB.

If the initial or subsequent contacts of the whisker to the post are too violent, then the whisker will be blunted to the extent that a hysteretic I-V curve may result. An entry of Y in the hysteretic column means that this has happened. Further discussion of this subject is given at the end of this section.

The output, P_{IF} , does not necessarily have a single maximum versus V . In fact, for low levels of microwave drive, there will be very sharp peaks in P_{IF} versus V [8]. However, as the drive power is increased these several maxima broaden until finally there is but a single very broad maximum. [Actually, a maximum for positive V and one for negative.] Sometimes the maximum output that could be obtained occurred before these maxima coalesced. In any case, the last column lists that value of V corresponding to the set of P_{drive} and

V values that maximized P_{IF} . All the P_{IF} data of Table I (and the corresponding V's) were taken for $N=3$. For this case, P_{IF} is much less sensitive to P_{K-band} than it is to P_{X-band} . To first-order, (for the $N = 3$ case) the optimum drive level (and this applies to both sources) is that which produces the maximum number of steps in the I-V curve. [It might be suggested that the smaller amount of P_{IF} that occurred with the low resistance junctions was just an impedance matching problem. This does not appear to be the case. A separate test indicates that the coaxial filter (fig. 2) can be adjusted to provide a good match to junctions ranging from 1 ohm to about 200 ohms. The filter was adjusted to maximize P_{IF} for each contact.]

The data of Table I were taken using six different whiskers. Three of the whiskers were made from solid niobium (Nb) wire of 0.005 cm (0.002 inches) diameter. One was made from solid Nb of 0.013 cm (0.005 inch) diameter. The other two were made from presharpended 0.005 cm (0.002 inch) diameter tungsten (W) wire which was subsequently plated with about 1 μ m of Nb. To my knowledge, the present work represents the first use of plated PCJJ's for multiplication and mixing. The solid Nb whiskers were pointed by etching in pure hydrofluoric acid.

An examination of the table shows that for every case in which P_{IF} was greater than 40 dB above the noise, the whisker used was either the 0.013 cm (0.005 inch) diameter solid Nb or of the Nb-plated W type. The probable reason for this is the extreme difficulty in making a junction with a large $I_0 R$ product and with I_0 of the order of 250 μ A when using the 0.005 cm (0.002 inch) Nb whisker. The usual problem is that, as a person attempts to get an acceptably large I_0 , the tip bends over and presumably such a large contact area results that the contact resistance becomes quite small and, at the same time, the supercurrent becomes excessively large.

A person intuitively feels that the I_0 , R and capacitance values of a given junction will depend upon whisker preparation (e.g., sharpness of point, state of oxidation) and upon the procedure of contacting the whisker to the post. The junctions that are available for study in a given experiment will then depend upon these fabrication conditions and their variability. As the discussion in items c and d of section IV imply, important things go on in dimensions that are of the order of the wavelength of visible light and smaller.

If one understood the microscopic details of the contact he might be able to explain why some contacts work well and others don't. In the absence of this knowledge one can only hope for a procedure that has a high probability of producing high quality junctions. During the course of the present work such a procedure was developed and is described in the appendix.

IV. DISCUSSION

With section III as background, it is now appropriate to reexamine points 1-4 of section I.

The data of Table I correlate well with the idea that I_0 needs to be of the order of 250 μ A to produce good P_{IF} . It might be suggested that the I_0 criterion and the $I_0 R$ criterion are not independent, that setting I_0 to about 250 μ A automatically gives a large $I_0 R$ product (The theoretical maximum $I_0 R$ for Nb is ≈ 2.4 mV). Table I seems to deny this because there are four contacts (#'s 16, 11, 20 and 7) that satisfy the I_0 criterion but whose resistance is low. It appears, therefore, that an $I_0 R$ product of the order of 1 mV or greater is an independent and important criterion.

With respect to the energy-gap, four of the 21 contacts exhibit energy gap structure in their I-V curves. Figure 1 is, in fact, the I-V curve of contact 18. Each of these four contacts produced P_{IF} of 53 dB or greater. The largest other output was 51 dB produced by contact 15 and it exhibited a "pathological" I-V curve. It is important to note that the energy gap criterion correlates very well with the I_0 and $I_0 R$ criteria discussed in the preceding paragraph.

Regarding steps in the I-V curve, the most surprising result is the strong negative correlation between the presence of microwave steps in the I-V curve and the level of P_{IF} . I do not have a quantitative explanation for this. There are, however, two qualitative things that can be said about factors that make steps difficult to see. First, noise will round the edges of the steps. The lower the frequency of the source which produces the steps, the closer spaced the steps will be. Close spacing in addition to noise rounding decreases the clarity of the steps. Second, for a given voltage spacing ΔV between steps, the current spacing between their centers, ΔI , will decrease with increasing resistance, i.e., $\Delta I = \Delta V/R$. If some mechanism prohibits the steps from being hysteretic, then their full height is limited to ΔI . Apparently such a mechanism exists because the only induced steps

(of which I am aware) that show hysteresis are associated with a closely coupled cavity resonance [19] which is not the case in the present work. Zimmerman and Sullivan have proposed that the absence of hysteresis in induced steps is due to noise which causes the junction to make a transition from the higher to the lower of the two possible energy states [20].

In the present experiment, steps were produced with the K-band source as well as the X-band. This was done by setting the attenuator for source #1 (the X-band source common to both the N=1 and N=3 experiments) to maximum. The switch was then put in position 3 and the junction irradiated with the K-band source. Any contact for which X-band steps could be produced also produced K-band steps. A contact that would not show X-band steps did not show K-band steps. The qualitative arguments given above probably bear on the unobservability of the steps but a quantitative explanation is lacking.

An important practical problem when a laser is used as the high frequency source is the difficulty in focusing upon the junction. In the past - when failing, after some effort at trying to get the PCJJ to respond to the laser - the quality of the junction has come into question. I presently believe that, if the I_0 , $I_0 R$, and energy gap criteria are all met, the junction is vindicated.

Consider now, items a-d of section I:

Item a: Spectrum spreading

The present experiment typically showed a 10-13 db decrease in maximum P_{IF} between the N=1 and N=3 cases. The frequencies involved preclude this fall-off being due to energy gap or capacitance effects (items b and d). The experiment was performed such that the junction properties remained essentially unchanged between the N=1 and N=3 cases for any given contact. The measured fall-off should therefore be a legitimate measure of the spreading of the supercurrent with increased N value. In the absence of a quantitative theory (based on the constant-current model) of P_{IF} versus the several parameters discussed in this paper, the only further thing to be said is that the optimum "local oscillator" power in the N=1 case was from 6.5 to 11 dB less than that for the N=3 case.

Item b: Energy gap

For Nb, the binding energy (proportional to the energy gap) of an electron pair corresponds to a

frequency of about 720 GHz. Since the highest drive frequency in this experiment was 25.1 GHz, energy gap effects should not have been a limitation. The laser experiments which are planned [2] will involve frequencies from ~ 600 GHz to 4.25 THz (4250 GHz). Presumably the energy gap effect will then become important. Whether it will be possible to sort out the effects, a-d, one from another remains to be seen.

The present experiment does, however, have a potentially important bearing on this effect. The success in getting high quality junctions by plating Nb on W probably means that other superconducting materials would also work. Thus, for example, it may be possible to use Nb_3Sn or Nb_3Ge and essentially double the cutoff frequency due to this effect.

Item c: Heating

The analysis of Tinkham et al [10] on high frequency limitations due to heating, contains a parameter, P_0 . This is essentially the level of ac power, absorbed by the junction, at which heating effects become important. They estimate the value of P_0 to be 10 μW . [This is based on adjusting the parameters of the theory to give a junction resistance (at large V) which is of the order of 10 ohms.] For the best contacts (17, 18, 21, 22), the local oscillator power at the top of the Josephson junction probe ranges from 36 μW (#22) to 5 μW (#18).

Since the characteristic impedance of the uniform part of the JJ probe is 50 ohms, the estimated values of absorbed power for #'s 22 and 18 are 0.16 μW and 0.50 μW respectively. I would, therefore, tentatively conclude that heating was not a significant limitation in this experiment. Estimating the drive level necessary to maximize P_{IF} for a laser frequency of 4.25 THz ($N \approx 460$), on the basis of the constant-voltage model [14], yields about 8 mW. It seems highly likely then, at the highest frequencies to which we intend to go, that heating will be a limitation.

Item d: Capacitance

Stewart has shown that when a junction is shunted by a sufficiently large capacitance, the resulting I-V curve will be hysteretic [21]. Others have realized this effect experimentally. A number of hysteretic I-V curves were obtained during the course of the present experiment. They are detrimental to obtaining good P_{IF} . Based on Stewart's work, an estimate has been made of the capacitance required to produce a hysteretic I-V curve typical of the present experiment. The result is $6.6 \times$

10^{-13} Farads. If it is assumed that this capacity results from a dielectric barrier between the post and the tip of the whisker (assuming $\epsilon = 25\epsilon_0$ - the approximate dielectric constant of niobium oxide - and a barrier thickness of 10^{-9} m = 10 Å) then the cross sectional dimension of the effective parallel-plate capacitor would be 1.7 μ m.

Microscopic examination of whiskers which have produced hysteretic I-V curves shows them to be blunted and with a flattened area whose cross sectional dimension is of the order of 2 μ m. It is essential to note, however, that on several occasions, whiskers which have produced hysteretic I-V curves have been recontacted and the resulting I-V curve has not been hysteretic. In addition, junctions have been formed which never showed hysteresis and upon subsequent examination there was essentially no evidence of blunting. I tend to agree with Zimmerman that the usual contact dimension is usually significantly smaller than the radius of curvature of the tip [11]. Although the above calculation gives an upper limit to the capacitance, I know of no way to calculate a believable lower limit. As additional feeling for the numbers involved, if the cross sectional dimension were 0.5 μ m then the capacitance would be 5.5×10^{-14} Farads and the resulting cutoff frequency would be about 0.6 THz.

An additional comment about high frequency limitations: The maximum P_{IF} obtained for the N=3 case was $\approx 3 \times 10^{-10}$ watts. If P_{IF} were proportional to N^{-2} , then the output power at 4.75 THz would be about 1×10^{-14} watts. Previous experience with essentially this same condition (extrapolated from N = 401 to N = 461) yielded about 2×10^{-15} watts [4]. There has been some experimental evidence for a fall-off in output going as N^{-2} and an analytical prediction of such behavior [4,7]. The data of Blaney et al suggests, however, that the fall-off is more rapid if the laser frequency is above ~ 1 THz [9]. In sum: the spectrum spreading effect almost surely increases monotonically with increasing N value. For large N and large values of 'laser' it is likely that heating will be important; energy gap effects probably will be important, and a significant degradation due to contact capacitance cannot be ruled out.

V. CONCLUSIONS

At microwave frequencies, there are three characteristics which, if simultaneously present, appear to assure a PCJJ which is a good frequency multiplier and

mixer. They are:

- (1) the critical current, I_0 , should be of the order of 250 μ A,
- (2) the undriven I-V curve should display "energy gap" structure,

and (3) the $I_0 R$ product should be of the order of 1 mV or greater.

The criterion of steps in the I-V curve, which is useful at laser frequencies, presents difficulties at microwave frequencies and this is discussed in the text.

The operational procedure - described in the appendix - of fabricating a PCJJ has a high probability of producing a good junction. In particular, the use of a tungsten whisker plated with a superconductor appears to be of significant importance in application.

There are four major effects that have potential for degrading multiplication and mixing performance at high frequencies. However, it is difficult to assess their absolute or even their relative effects at frequencies above about 1 THz. The present paper does, however, give quantitative information about spectrum spreading in going between the N=1 and N=3 cases at microwave frequencies.

ACKNOWLEDGEMENTS

Several people at NBS have been helpful in the work reported here. I am indebted to C.A. Hamilton for the use of his PCJJ mount and to L.O. Mullen for the fabrication of the Nb-plated tungsten whiskers. Discussions about Josephson junctions with J.E. Zimmerman were helpful. Useful comments on subject matter were made by R.M. Garvey, S.R. Stein and H. Hellwig.

APPENDIX

During the course of this work, a procedure has evolved which I have found to be very useful in producing a useable contact. In this procedure, one lead of an ordinary ohmmeter (multimeter) is connected to the post and the other to the whisker. The ohmmeter is set to the x100 scale and, if the contact is carefully made, the initial contact resistance - as indicated by the meter - is usually a few hundred ohms or more. The supercurrent at that time is almost always unacceptably small (less than 50 μ A). The next step is not to further increase the pressure upon the contact (which will usually blunt the point) but rather to change the scale of the ohmmeter to x10 and, perhaps, to x1. (On the

x100 scale I try and get a junction to stabilize at about 200 ohms. A 200 ohm resistor draws 2.3 mA on this scale. Subsequent switching of a PCJJ down to the x10 scale usually results in a resistance of 20 - 30 ohms. A 26 ohm resistor draws 19 mA on this scale. Switching to x1 usually results in a 3-5 ohm resistance. A 4.4 ohm resistor draws 140 mA.] This often will result in a supercurrent of about the right magnitude. If the I_0 is a little large or if the shape of the I-V curve needs improvement (in terms of the I_0 and energy gap criteria) then it is often possible to improve the junction by slightly backing off the pressure on the contact. Sometimes it is necessary to completely open the contact and begin the procedure again. In my experience this procedure has worked much more than half the time for the plated whiskers and the larger diameter whiskers. On the other hand, the often-used procedure of closing the contact with little or no bias across it has a high probability of damaging the tip beyond usefulness.

REFERENCES

- [1] S.R. Stein, Allan S. Risley, H. Van de Stadt and F. Strumia, "High Speed Frequency Modulation of Far Infrared Lasers using the Stark Effect," to be published in *Applied Optics*, July 1977.
- [2] S.R. Stein and H. Van de Stadt, "Electronic Tuning and Phase-Lock Techniques for Far Infrared Lasers," to be published in *Proc. 31st Annual Symposium on Frequency Control*, Atlantic City, NJ, June 1977.
- [3] S.R. Stein and J.P. Turneaure, "Superconducting Cavity -Stabilized Oscillators with Improved Frequency Stability," *Proc. IEEE* 63, pp. 1249 (1975).
- [4] D.G. McDonald, A.S. Risley, J.D. Cupp, K.M. Evenson and J.R. Ashley, "Four hundredth order harmonic mixing of microwave and infrared laser radiation using a Josephson junction and a maser," *Appl. Phys. Lett.*, Vol. 20, pp. 296-299, April 1972.
- [5] T.G. Blaney and D.J.E. Knight, "Direct 825th harmonic mixing of a 1 GHz source with an HCN laser in a Josephson junction," *J. Phys. D: Appl. Phys.*, Vol. 7, pp. 1882-1886, 1974.
- [6] D.G. McDonald, A.S. Risley, J.D. Cupp and K.M. Evenson, "Harmonic mixing of microwave and far-infrared laser radiation using a Josephson junction," *Appl. Phys. Lett.*, Vol. 18, pp. 182-184, February 1971.
- [7] T.G. Blaney and D.J.E. Knight, "Experiments using a superconducting point-contact harmonic mixer near 1 THz," *J. Phys. D: Appl. Phys.*, Vol. 6, pp. 936-952, 1973.
- [8] A.S. Risley, E.G. Johnson, Jr. and C.A. Hamilton, "Analog computer studies of frequency multiplication and mixing with a Josephson junction," *Proc. 1976 Applied Superconductivity Conference*, Stanford University, August 1976, pp. 381-384.
- [9] T.G. Blaney, N.R. Cross and D.J.E. Knight, "Harmonic mixing and frequency measurement at 2.5 THz using Josephson junctions," (has been submitted to: *J. Phys. D: Appl. Phys.*).
- [10] M. Tinkham, M. Octavio and W.J. Skocopol, "Heating effects in high frequency metallic Josephson devices: voltage limit, bolometric mixing and noise," *J. Appl. Phys.*, Vol. 48, pp. 1311-1320, March 1977.
- [11] J.E. Zimmerman, "A review of the properties and applications of superconducting point contacts," *Proc. 1972 Applied Superconductivity Conference*, Annapolis, MD, May 1972, IEEE, Inc. NY (1972).
- [12] C.A. Hamilton, "Analog simulation of a Josephson junction," *Rev. Sci. Instr.*, Vol. 43, No. 3, pp. 445-447, March 1972.
- [13] V.E. Kose and D.B. Sullivan, "Influence of external noise on microwave-induced Josephson steps," *J. Appl. Phys.*, Vol. 41, No. 1, pp. 169-174, January 1970.
- [14] C.C. Grimes and Sidney Shapiro, "Millimeter-wave mixing with Josephson junctions," *Phys. Rev.*, Vol. 169, No. 2, pp. 397-406, May 1968.
- [15] T.G. Blaney, "Heterodyne laser radiation detection at 891 GHz using Josephson point contacts," *Rev. Phys. Appl.*, Vol. 9, pp. 279-284, 1974.
- [16] T.G. Blaney and D.J.E. Knight, "Heterodyne reception at 891 GHz using a Josephson harmonic mixer and a microwave local oscillator," *J. Phys. D: Appl. Phys.*, Vol. 7, pp. 1887-1893, 1974.
- [17] Y. Taur, J.H. Claassen and P.L. Richards, "Josephson junctions as heterodyne detectors," *IEEE Trans. MTT*, Vol. 22, No. 12, pp. 1005-1009, Dec. 1974.
- [18] Private communication, J.E. Zimmerman and D.B. Sullivan.
- [19] Arnold H. Silver and James E. Zimmerman, "Josephson weak-link devices," *Applied Superconductivity*, Vol. 1, Editor Vernon Newhouse (Academic Press, 1975).
- [20] J.E. Zimmerman and D.B. Sullivan, "High frequency limitations of the double-junction SQUID amplifier," to be published.
- [21] W.C. Stewart, "Current-voltage characteristics of Josephson junctions," *Appl. Phys. Lett.*, Vol. 12, No. 8, April 1968.

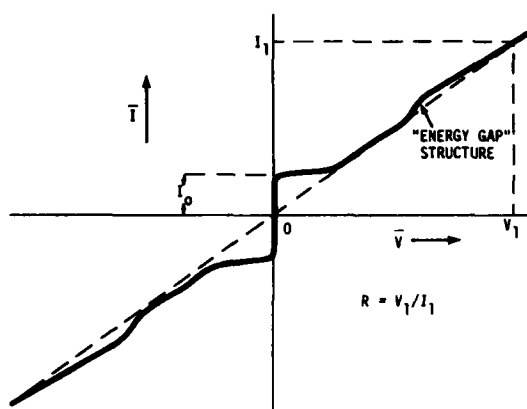


FIGURE 1. The total dc current, I_t , through the PCJJ is plotted versus the dc voltage, V , across it. For large values of V (say, V_1) I_t asymptotically approaches the dashed line.

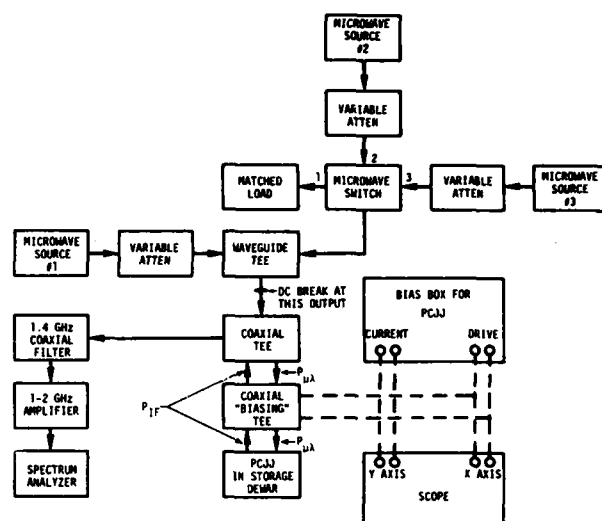


FIGURE 2. Block diagram of experimental set-up.

EXTENT OF STEPS (μV)	$2 I_0$ (μA)	R (ohms)	$I_0 R$ (mV)	CONTACT #	P_{IF} FOR $N = 3$ (dB ABOVE NOISE IN 30 KHz B.W)	HYSTERETIC ?	\bar{V} FOR MAX P_{IF} (μV)	TYPE OF WHISKER
1,000	580	N.M. (NOT MEASURED)	--	10	N.M.	Y		+
800	3,850	< 1	--	23	31	Y	10	†
> 700	680	2.8	0.95	16	41		30	*
~ 700	1,050	< 1	--	3	N.M.			+
700	800	< 1	--	11	N.M.	Y		+
700	1,500	< 1	--	14	N.M.			*
600	1,400	< 1	--	12	N.M.	Y		+
~ 500	180	< 1	--	2	N.M.			+
450	530	~ 1	~ 0.5	20	34		10	*
420	1,600	< 1	--	8	35		4	+
180	520	< 1	--	7	35		45	+
NOT VISIBLE	480	~ 9	2.2	21	54		500	+
	300	8.8	1.3	18	59		600	*
	300	8.4	1.3	19	41	Y	200	*
	150	~ 7	0.53	5	N.M.	Y		+
	560	6	~ 1.7	22	53		100	†
	200	4.5	0.45	1	40	Y	550 & 90	+
	400	4.2	0.84	17	57		500	*
	320	4.0	0.64	24	47		250	†
	140	2.0	0.14	4	N.M.			+
	50	0.9	0.045	6	25		10	+
	1,800	< 1	--	15	51		N.M.	*

+ = 0.002" SOLID Nb

† = 0.005" SOLID Nb

* = 0.002" Nb-PLATED W

OPTICAL ELECTRONICS, EXTENSION OF MICROWAVE TECHNIQUES INTO THE OPTICAL REGION

C. F. Davis, Jr., G. Elchinger, A. Sanchez, K. C. Liu and A. Javan

Massachusetts Institute of Technology

Infrared diode work began some ten years ago with point contact MOM junctions. These were used as harmonic generators and mixers, often of fairly high order. They have even been used in stabilized harmonic generator chains through $2\text{ }\mu\text{m}$.¹ Such devices can form the basis of active far infrared (FIR) elements and circuits for reradiation of laser sidebands, signal processing in this region of the spectrum and studying extremely short pulse phenomena. Recently work has been done on MOM devices evaporated through masks or photolithographically defined. I will show you why small area devices are essential to extending the operating frequency and how such devices can perform active circuit functions in the FIR and perhaps someday on into the visible; already MOM diodes are responding non-thermally to visible radiation. A study² are just completing shows that a point contact MOM behaves according to classical antenna theory in the FIR. Some of our latest work is described on printing MOM circuits for a real time holographic array.

The MOM devices operate by nonlinearities in their conductance, which must be the order of the signal source (the antenna at about $100\text{ }\Omega$). This junction is then shunted by its own capacitance which restricts its high frequency operation. Junction conductance decreases exponentially with thickness while capacitance goes as the reciprocal; hence the thinnest junction must have the most favorable conductance capacitance ratio. A junction with barrier thickness about the typical $10\text{ }\text{\AA}$ would have resistance of $10^2\text{ }\Omega/\text{A}$ (A in μm^2) requiring about $1\text{ }\mu\text{m}^2$ to match the typical antenna resistance. Orders of magnitude change in the area can be compensated for by small changes in barrier thickness to maintain junction impedance comparable to the antenna resistance ($100\text{ }\Omega$) while adjusting the capacitance. The response time would then be given by $RC = 10^{-12}\text{ A}$. (A in μm^2)

We have considered several FIR functional circuits; these have not yet been built. However, fabrication would require two successive photolithographically defined metalizations, the first capable of forming and supporting a stable barrier layer about $10\text{ }\text{\AA}$ thick and the second of adhering to the substrate and making good contact to the barrier layer. One circuit usable as a parametric subharmonic generator might consist of a circle half made of each metal, overlapping at one intersection to form a small area MOM diode and at the other a large area ohmic contact. Inherent inductance and capacitance of this circuit and junction can make it resonant in the FIR. With appropriate input driving signal and bias leads and a realizable junction nonlinearity, a Q of 5 would be adequate to cause oscillation at half the signal frequency.

Some particularly interesting work has been with high speed deposited negative resistance junctions formed by evaporating a narrow lead stripe on tin with a thin surface oxidation. This junction is cooled to about 2°K where both metals are superconducting and irradiated with a focused argon laser

beam; definite non-thermal response was observed. Responses of lead on aluminum junctions were observed, again at 2°K (aluminum above its superconducting transition). The response to X-band and $5145\text{ }\text{\AA}$ argon laser radiation was very similar showing peaks in exactly the same locations. The principal structural response occurs at dc bias voltage less than 20 mV and is due to the superconducting transition and photon scattering in the lead film.

A theoretical analysis² of the MOM antenna/diode as a detector of microwave and infrared radiation is being conducted and FIR experiments evaluated to examine the consistency of the theory. The antenna is coupled directly into the diode. An equivalent circuit is used to represent the system of the antenna and its coupling to the diode. A Stratton³ tunneling model represents the nonlinear character of the junction. Detector performance is shown to obey experimentally verified laws and determine an optimum junction thickness and area for each frequency. It is shown that the detectivity at room temperature can be as high as $10^{10}\text{ watts}^{-1}\text{ Hz}^{-1/2}$ at frequencies of 10^{14} Hz in the infrared. Experimental results show that for small focusing angles, $\theta_f = D/f$ (where $f = 12.7\text{ cm}$ in this case), the efficiency η is proportional to θ_f^2 (see Fig. 1) consistent with the concept of effective aperture. The proportionality constant (at $337\text{ }\mu\text{m}$ wavelength) agrees within a factor of two with that expected from our theory; this discrepancy may be caused by uncertainty in the calorimetric laser power measurement. As the focusing angle is increased to the width of the major radiation lobe, the coupling efficiency saturates to about 3%, in agreement again with the theory. Different antenna lengths give proportionally different coupling efficiencies for small focusing angles and the same saturated value for larger angles.

To confirm that the FIR detection arises from the same mechanism as that for rf detection (i.e. rectification), we checked, using a balancing technique, that the same detected voltage comes from an rf signal as from a preadjusted ir signal, as the bias is increased up to about 100 mV .

This balancing technique has been applied to the study of the type of printed diodes integrated to an antenna described in Ref. 4. Preliminary results show that some structures have coupling efficiencies as high as 2 or 3%, comparable to that of the mechanical point contact diode. This technique is applicable regardless of the nature of the rectification mechanism as long as it is the same at rf and ir frequencies. For example, we have also fabricated antenna structures where the detecting element is an evaporated micron-sized, thin-film wire without a junction. A weak bolometric response, with the same bias dependence, was observed at both frequencies and the best coupling efficiency was found to be around 1%.

We have studied the behavior of the point contact diode nonlinearities by observing the bias dependence of the rectified signal when a known

amount of rf power is coupled to the diode. These results can be compared with the predictions of the tunneling model. Values for the barrier parameters are found to be reasonable.

The diode's I-V characteristic up to the third order can be expressed as:

$$I = \frac{1}{R} (V + mV^2 + nV^3)$$

Then the rectified voltage, V_r , when an rf voltage $V_D \cos \omega t$ is coupled to these diode nonlinearities is given by

$$V_r = (m + 3n V_b) \frac{V_D^2}{2}$$

The experimental dependence of V_r on V_D shown in Fig. 2, can be approximated by a straight line in the low bias region. We can then obtain for the 2 K-ohm diode, $m = .16 \text{ V}^{-1}$ and $n = 1.8 \text{ V}^{-2}$.

These values for m and n , which are independent of the contact area, are in general agreement with those obtained from the model assuming reasonable values for the barrier parameters. From the tunneling model, we can express the average potential, ϕ_0 , and thickness of barrier, L , as:

$$\phi_0 = \frac{1}{4\sqrt{6n}} \ln \frac{324 R_D a}{4\sqrt{6n}}$$

$$L = \frac{4\sqrt{6n\phi_0}}{1.025}$$

By estimating the value of contact area as $a = .1 \mu\text{m}^2$, we can then obtain $\phi_0 = .65 \text{ eV}$ and $L = 10.3 \text{ \AA}$ which are quite reasonable and change only logarithmically with the estimated area. Furthermore, the value of the asymmetry factor, α , according to the theory is given by

$$\alpha = m\sqrt{6/n}$$

which, in the case of Fig. 2, is equal to 0.29 and does not depend on the estimated area.

As the resistance of the diode is lowered by adjusting the pressure of contact, the value for n decreased monotonically. For example; if $R_D = 50 \text{ ohms}$ we get $n = 1.39$ so that $\phi_0 = 0.43 \text{ eV}$ and $L = 7.3 \text{ \AA}$. At these low values of L and ϕ_0 , the W.K.B. approximation used in calculating the tunneling probability begins to lose its applicability.

We have shown that in agreement with the theory, the zero bias responsivity, $\beta_1 (=m)$, remains nearly constant as the resistance is varied over two orders of magnitude. By studying the laser rectification as the resistance of the diode, R_D , "relaxes" continuously and slowly (presumably keeping the same asymmetry factor α) from 10 to few thousand ohms, we were able to fit the result to the expression:

$$V_r = V_0 \left(\frac{R_D}{R_D + R_A} \right)^2$$

with two adjustable parameters, V_0 and R_A (antenna

resistance).

A 2 x 2 array of antenna/diodes has been fabricated for ultimate use as a real time holographic imaging array. This is now being evaluated as individual infrared rectifier/mixers. In operation, two infrared beams differing by a microwave frequency will irradiate the array. From the amplified difference frequency we can obtain relative phase information of the two signals. This provides the information required to define and construct a real time hologram.

1. D. A. Jennings, F. R. Peterson, K. M. Evenson, Appl. Phys. Letts. 26, 510 (1975).
2. To be published.
3. R. Stratton, J. Phys. Chem. Solids 23, 1177 (1962).
4. J. Small, G. M. Elchinger, A. Javan, A. Sanchez, F. L. Bachner, D. L. Smythe, Appl. Phys. Letts. 24, 275 (1974).

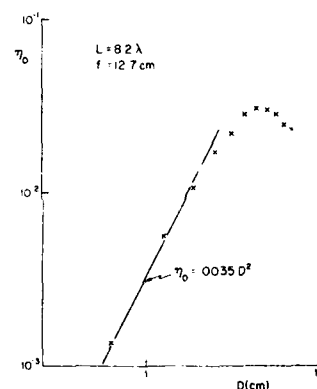


Figure 1. The antenna/diode efficiency is plotted as a function of iris diameter, D . It should be noted that saturation at high values of D comes from a combination of approaching the beam diameter and exceeding the width of the first lobe. The proportionality constant, .0035, is within a factor of two of the theoretical value.

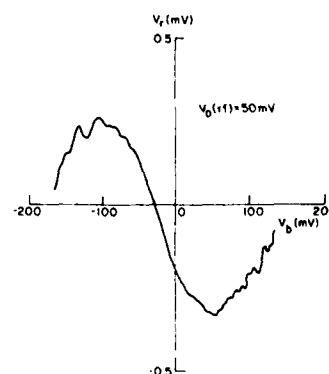


Figure 2. The experimental dependence of V_r on V_D is presented. Diode resistance was measured as 2 K Ω and values of $m = .16 \text{ V}^{-1}$ and $n = 1.81 \text{ V}^{-2}$ were calculated from this plot.

PROGRESS IN CO₂ LASER STABILIZATION*

Charles Freed
Massachusetts Institute of Technology
Lincoln Laboratory
Lexington, Massachusetts 02173

Summary

A family of stable CO₂ lasers have been under development at M.I.T., Lincoln Laboratory for the past few years. These lasers are used in a variety of applications and under different operating conditions.

Stability data have been obtained for free-running, offset-locked and line-center stabilized lasers. Measurements of spectral purity (frequency domain), and of the fractional frequency stability (time domain) of the laser beat notes will be presented.

A two-channel, line-center stabilized CO₂ laser heterodyne frequency calibration system will also be described, together with the latest data obtained for the various CO₂ isotope laser transitions. The new data are 10³ to 10⁶ times more accurate than previously published results and will provide secondary frequency references over a wide portion of the infrared spectrum.

Key Words: Stable CO₂ lasers, short-term stability, long-term stability, pressure shift, CO₂ isotope constants.

Introduction

This paper reviews the stability obtained with various CO₂ lasers and under different operating conditions. New results are emphasized while previously published details on laser design, experimental procedure and long-term stabilization are left to the references.

In the first part of the paper results obtained with free-running lasers are described. The second part concentrates on line-center locked CO₂ lasers and some of their applications.

Open Loop Stability

In this section we shall first describe lasers and results applicable to high resolution optical radars. Figure 1 shows the basic laser structure which was previously described in greater detail.^{1,2} Figure 2 illustrates the real time power spectrum of the beat note between two free running lasers, similar to the one shown in Figure 1. This figure and the conditions under which it was obtained was also published previously² and is repeated here for completeness sake. Figure 3 shows a real time power spectrum similar to the one in Figure 2, but measured under somewhat noisier environmental conditions.

The mostly discrete modulation sidebands in Figure 3 are primarily due to line frequency harmonics, structure vibrations, acoustic noise and optical feedback from the detector. Note, that the modulation sidebands in Figure 3 may be enclosed under a Gaussian envelope with $\sigma \approx 209$ Hz. Although laser stabilities

are most conveniently measured by heterodyning two lasers, the results are not altogether foolproof because the disturbances causing frequency jitter of the lasers may be at least partially correlated. In an optical radar one may compare the laser with its own output delayed by the round trip time to and from the target. Hence, effects due to disturbances with correlation times less than the round trip time of the transmitted signal will be included in the measured beat note spectrum.

Figure 4 shows a block diagram of a 10.6 μ m laser radar at the Lincoln Laboratory Firepond Facility^{3,4,5,6} in Westford, MA. In Figure 4 wavy and solid lines denote optical and electrical signal paths, respectively. The 0.5 meter local and the 1.5 meter master oscillators are similar to the one shown in Figure 1. The higher power master oscillator is offset-locked to the local oscillator using a stable 10 MHz frequency discriminator and a frequency control servo loop with unity gain at 3 kHz. Figure 5 shows the real time power spectrum of the 10 MHz beat note of the two offset-locked lasers. The analyzer resolution in Figure 5 is set to 0.02 Hz; thus the observed narrowness of the beat note spectrum indicates that the master oscillator tracked the local oscillator within less than one part in 10¹⁵ during the 50 sec observation period required to obtain a high resolution measurement such as shown in Figure 5.

The 10.6 μ m laser radar shown in Figure 4 has been used to make observations on GEOS-III, a NASA geodetic satellite equipped with an IRTAN II solid cube corner retro-reflector. Radar returns from GEOS-III have been used to determine the radial velocity of the satellite using Doppler measurements, and to set an upper bound to the laser oscillator instability.

A logarithmic display of the power spectra of a consecutive sequence of radar return signals is shown in Figure 6. Table I summarizes the applicable operating conditions.

The returns in Figure 6 were obtained near the closest approach of the satellite, corresponding to a Doppler shift of 64 MHz. At this point of the orbit the Doppler frequency was changing by 10 MHz/sec. This change in frequency amounted to a 40 kHz shift during the 0.004 sec radar transmission. This frequency shift, together with the main Doppler shift was removed by orbit-fitting in the computerized processing of the data. Thus the individual spectra of the radar returns shown in Figure 6 should ideally approach a $(\frac{\sin x}{x})^2$ distribution corresponding to the Fourier transform of the 10.6 μ m single frequency 0.004 sec radar pulse. The spectral width one may observe on the returns in Figure 6 is approximately twice the theoretical 500 Hz that should occur between the first minima. The excess width and smearing of the return spectra is due to instabilities and jitter in the entire, rather complex system, including the laser

* This work was sponsored by the Advanced Research Projects Agency of the Department of Defense and also by the U. S. Energy Research and Development Administration.

oscillators. Needless to say, the system performance is steadily improving as the operating conditions and components giving rise to instabilities are gradually corrected; however, even the present level of performance corresponds to an orbital velocity precision of 1.25 mm/sec, and one may put an upper bound of about one part in 10^{11} on the short-term stability of the laser oscillators.

Open-Loop Stability Of Microwave Self-Beats

The results in the remainder of this paper were obtained with grating controlled lasers, which have been previously described.^{2,7} Such a grating controlled laser is shown in Figure 7.

Microwave (or lower) frequency generation may be achieved with a single laser filled with a mixture of CO_2 isotopes. The spectral purity and long-term stability of the (self-) beat notes obtained in this way may only be compared to stabilized oscillators of the highest quality. Figure 8 shows the spectrum analyzer display of the 3165 MHz self-beat of the $^{14}\text{C}^{16}\text{O}_2$ $00^0_1-[10^0_0, 02^0_0]_{II}$ P(10) and the $^{12}\text{C}^{16}\text{O}_2$ $00^0_1-[10^0_0, 02^0_0]_{II}$ R(18) transitions. Time domain frequency stability measurements of the same 3165 MHz beat note are shown in Figure 9. In order to obtain the fractional frequency stability, the measured Allan Variance was divided by $\sqrt{2}$ times the laser frequency. Table II gives the sample size, m , for each observation time, τ , used to obtain the Allan Variance.

Figure 10 shows an even better frequency stability measurement of the 593 MHz beat note between the $^{14}\text{C}^{16}\text{O}_2$ $00^0_1-[10^0_0, 02^0_0]_{II}$ P(38) and the $^{12}\text{C}^{16}\text{O}_2$ $00^0_1-[10^0_0, 02^0_0]_{II}$ P(16) laser transitions. Each data point in Figures 9 and 10 is based on an independent set of consecutive m samples. Table III shows the sample size, m , for each observation time, τ , of Figure 10.

The spectral purity of Figures 8, 9 and 10 is explained by the fact that the fractional frequency stability at the microwave beat frequency will be identical to the fractional stability of the laser frequency itself. It should be clear by now that short-term stabilities of 10^{-11} to 10^{-13} may be routinely achieved with well designed and acoustically shielded CO_2 lasers.

Two additional facts are noteworthy in Figures 9 and 10. The frequency stability gets better for shorter observation times, a very desirable requirement in radar-like applications. As a matter of fact, inadequate frequency resolution of our present equipment prevented meaningful measurements of the Allan Variance for observation times less than 1 sec in cases like the ones illustrated by Figures 9 and 10. A second noteworthy feature of such self-beats is that they may be utilized for direct comparison of optical and microwave frequency domains.

Long-Term Stability and Some Of Its Applications

It was previously shown^{8,9} that CO_2 lasers can be frequency stabilized by using the standing-wave saturation resonances in a low-pressure, room-temperature, pure CO_2 absorber via the intensity changes observed in the collisionally-coupled spontaneous emission band at 4.3 μm . More recently, significant improvement in signal-to-noise ratios with new, low pressure CO_2 stabilization cells external to the lasers was also demonstrated¹⁰.

The experimental results in the remaining part of this paper were obtained with the two-channel line-

center stabilized CO_2 isotope calibration system which was previously described⁷. Figure 11 shows the block diagram of this system.

In order to investigate in the most direct fashion the various parameters affecting long-term frequency stability, we carried out extensive measurements of the Allan Variance on the 2,697.86 MHz beat frequency between the $^{12}\text{C}^{16}\text{O}_2$ $00^0_1-[10^0_0, 02^0_0]_{II}$ band P(20) and the $^{13}\text{C}^{18}\text{O}_2$ $00^0_1-[10^0_0, 02^0_0]_{II}$ band R(24) laser transitions⁷. By using two different CO_2 isotope lines, frequency pulling due to optical feedback was reduced and the 2,698 MHz beat frequency output of the HgCdTe photodiode was directly measured by a microwave frequency counter; thus only two independently lockable lasers and a single microwave frequency counter were utilized in the experiments to be described in this section (note that Figure 11 indicates a separate microwave local oscillator and a second counter; neither of these were necessary for the stability measurements).

Since each laser was assumed to contribute equally to the instability, the measured Allan Variance was divided by $\sqrt{2}$ times the laser frequency (2.8306×10^{13} Hz) in order to derive the fractional frequency stability for a single laser as a function of sample time, τ , (gating time of the frequency counter).

The stability of the lasers may be best summarized by Figure 12 where each circle or cross represents an Allan Variance measurement based on a sample size, m , as specified in Table IV.

The fractional stability of the beat note of the two lasers under free running conditions is denoted by crosses and may be reasonably well approximated by σ_y (unlocked) $\approx 10^{-10} \times \tau^{0.72}$ which of course indicates the drift rate of the lasers relative to each other.

The circles represent the results obtained with each laser independently locked to its own reference absorption cell filled with 40 mTorr of $^{12}\text{C}^{16}\text{O}_2$ and $^{13}\text{C}^{18}\text{O}_2$, respectively.

As Figure 12 indicates, there were three consecutive sets of measurements made, each based on m samples for any given observation time, τ . The locked laser stability may be described by σ_y (locked) $\approx 6 \times 10^{-12} \times \tau^{-1/2}$.

Thus, the instability of the lasers became less than 1×10^{-12} for sample times $\tau \geq 40$ seconds. A fractional stability of 10^{-12} corresponds to $\sigma_y \approx 28$ Hz fluctuation in the laser frequency. Since the piezoelectric mirror tuning rate is about 200 kHz/volt, the phase sensitive detector output stability (drift) must be less than 150 μV to achieve 1×10^{-12} long-term stability. Such low, long-term drift was clearly beyond the capability of the ten year old lock-in amplifiers used to obtain these results; therefore, the measurement of longer term stability with larger sample sizes was not seriously pursued in the current phase of our experiments. However, the stability we did obtain was quite sufficient to carry out accurate and reproducible measurements of pressure shift in CO_2 .

In the pressure shift experiments both lasers are locked to their individual reference cells, and the shift in the beat frequency is measured as a function of pressure change in one of the cells, with the pressure held constant in the second cell in order to obtain a stable reference laser frequency.

Figure 13 shows the last two digits of the 2,697.862 ± 6 kHz beat frequency of the $^{12}\text{C}^{16}\text{O}_2$ laser

$00^01-[10^00, 02^00]_I$ band P(20) and the $^{13}\text{C}^{18}\text{O}_2$ laser $00^01-[10^00, 02^00]_I$ band R(24) transitions as a function of pressure in the $^{12}\text{C}^{16}\text{O}_2$ reference cell. Each circle in Figure 13 is based on an observation time of $\tau = 10$ seconds and a sample size of $m = 26$. As Figure 13 indicates, two independent, consecutive sets of measurements give results which are barely distinguishable from each other and are within one $\sigma_y(\tau)$ of the solid line. A straight line fitting (by the least-squares method) of the data between 0 and 60 mTorr yields a -108.6 Hz/mTorr frequency change with increasing pressure in the $^{12}\text{C}^{16}\text{O}_2$ reference cell (red shift). The pressure shift becomes larger and nonlinear at higher pressures, but for the purpose of frequency stabilization only the low pressure region is of interest. It is rather interesting to note that our original 1970 estimate^{2,8} was also about 100 Hz/mTorr red shift for the same P(20) laser transition.

Analogous data for the $^{13}\text{C}^{18}\text{O}_2$ $00^01-[10^00, 02^00]_I$ band R(24) transition gave a 168.6 Hz/mTorr red shift in the 0 - 60 mTorr pressure range. Neither repeated breaking and resetting of the frequency lock, nor refilling of the lasers and the reference cells altered the results shown in Figures 12 and 13.

In the final portion of this paper new results relating to CO_2 isotope lasers will be given. We have previously reported^{11,12} on the determination of transition frequencies, band centers and rotational constants of $^{12}\text{C}^{18}\text{O}_2$, $^{13}\text{C}^{16}\text{O}_2$ and $^{13}\text{C}^{18}\text{O}_2$ isotope lasers with accuracies of about 3 MHz.

Optical heterodyne techniques were used to generate beat frequencies between two lasers filled with different CO_2 isotopes. The rotational constants are computed by fitting the measured difference frequencies in a least-squares sense to the expansion of the line frequencies

$$\begin{aligned} \nu = & \nu_0 + B_u [J'(J'+1) - J(J+1)] - (B_L - B_u) J(J+1) \\ & - D_u [J'^2(J'+1)^2 - J(J+1)^2] + (D_L - D_u) J^2(J+1)^2 \\ & + H_u [J'^3(J'+1)^3 - J^3(J+1)^3] - (H_L - H_u) J^3(J+1)^3 \\ & - L_u \dots \end{aligned} \quad (1)$$

where $J' = J-1$ for a P(J) line and $J' = J+1$ for an R(J) line.

The experimental procedure is similar in principle to the one previously described^{12,13}, and this paper will only list the latest results which are summarized in Table V.

The band centers and rotational constants of $^{12}\text{C}^{16}\text{O}_2$ listed in Table V were originally determined by Petersen et al.¹³, and were used as fixed references in the computation of vibrational - rotational constants of the other CO_2 isotopes in Table V. Upon completion of our data taking in the near future, we shall relax this restriction, and will simultaneously determine the rotational constants of all the isotopes since our measurements include much higher J-transitions than reported previously¹³. Thus the constants listed in Table V should not be considered as the final results of our isotope measurements; however, they do represent an approximately 10^3 - fold improvement over our previously published data for $^{13}\text{C}^{16}\text{O}_2$ and $^{13}\text{C}^{18}\text{O}_2$ and an even greater improvement for $^{14}\text{C}^{16}\text{O}_2$. Figure 14 graphically illustrates the lasing transitions of the CO_2 isotopes listed in Table V and the spectral region they occupy.

Conclusion

Even the few examples given in this paper make it quite clear that the spectral purity, frequency stability and resettability, together with the availability of well over a thousand lasing transitions uniquely endow the CO_2 system for direct use in high resolution spectroscopy, or as secondary frequency standard in heterodyne spectroscopy with tunable lasers, or in precision IR synthesis which involves frequency mixing.

A systematic and precise evaluation of the band centers, rotational constants, and lasing transition frequencies of the CO_2 isotopes is under way. These data will also be of great value in evaluating the potential function under the influence of which the nuclei are moving.

Acknowledgments

The author is deeply appreciative to R. G. O'Donnell for assistance in all phases of the stability and isotope measurements and to many other colleagues who contributed to the design of the 10.6 μm radar and who provided the data relating to the GEOS-III experiments.

References

1. C. Freed, "Design and Short-Term Stability of Single-Frequency CO_2 Lasers, IEEE J. Quan. Elect. QE-4, 404-408 (June 1968).
2. C. Freed, "Designs and Experiments Relating to Stable Lasers, "Proceedings of the Frequency Standards and Metrology Seminar, University Laval, Quebec, Canada, 226-261 (1 September 1971).
3. R. H. Kingston and L. J. Sullivan, "Coherent Infrared Radar", SPIE 69, 10-13 (1975).
4. T. R. Gurski, "Laser Frequency Stability Limitations On A 10.6 μm Laser Radar", paper presented at OSA, Tucson, Arizona (18-22 October 1976).
5. R. H. Kingston, "Coherent Optical Radar" (1977) to be published.
6. R. Teoste, W. J. Scouler and D. L. Spears, "Coherent Monopulse Tracking With a 10.6 μm Radar", Paper 9.5 presented at the CLEA Conference in Washington, D. C. (1-3 June 1977).
7. C. Freed, R. G. O'Donnell and A. H. M. Ross, "Absolute Frequency Calibration of the CO_2 Isotope Laser Transitions", IEEE Trans. Instrum. Meas. IM-25, 431-437 (December 1976).
8. C. Freed and A. Javan, "Standing-Wave Saturation Resonances in the 10.6 μm Transitions Observed In A Low-Pressure Room-Temperature Absorber Gas", Appl. Phys. Lett. 17, 53-56 (15 July 1970).
9. C. Freed and A. Javan, "Standing-Wave Saturation Resonances in Room Temperature CO_2 10.6 μm Absorption Lines", Paper 4.4 presented at the 1970 Sixth International Quan. Elect. Conference, Kyoto, Japan (September 1970).
10. C. Freed, "Frequency Stabilization of CO_2 Lasers", Proc. of the 29th Annual Symposium on Frequency Control, Atlantic City, NJ, 330-337 (28-30 May 1975).
11. C. Freed, D. L. Spears, R. G. O'Donnell,

and A. H. M. Ross, "Precision Heterodyne Calibration", Proc. of the Laser Spectroscopy Conference, Vail, Colorado (25-29 June 1973). Also in Laser Spectroscopy, 171-191, Plenum Press, 1975 (R. A. Brewer and A. Mooradian, Editors).

12. C. Freed, A. H. M. Ross and R. G. O'Donnell, "Determination of Laser Line Frequencies and Vibrational-Rotational Constants of the $^{12}\text{C}^{18}\text{O}_2$, $^{13}\text{C}^{16}\text{O}_2$ and $^{13}\text{C}^{18}\text{O}_2$ Isotopes from Measurements of CW Beat Frequencies with Fast HgCdTe Photodiodes and Microwave Frequency Counters", J. Molecular Spectrosc. 49, 439 (1974).
13. F. R. Petersen, D. G. McDonald, J. D. Cupp and B. L. Danielson, "Accurate Rotational Constants, Frequencies, and Wavelengths from $^{12}\text{C}^{16}\text{O}_2$ Lasers Stabilized by Saturated Absorption", Proc. of the Laser Spectroscopy Conference, Vail, Colorado (25-29 June 1973).

10.6 μm LASER RADAR OBSERVATION OF GEOS-III SATELLITE

RANGE:	1007 km
DOPPLER VELOCITY:	- 318 m/sec
ELEVATION:	56.11°
PULSE DURATION:	0.004 sec
PULSE REPETITION RATE:	62.5/sec (0.016 sec)
FULL SPECTRAL WIDTH OF SIGNAL RETURNS:	$\Delta\nu = 693 (\pm 99)$ Hz 10 db BELOW PEAK
AVERAGE PULSE TO PULSE JITTER:	209 Hz

TABLE I

τ (seconds)	1	2	4	5	7	10
m	101	51	51	51	26	26

Sample size, m, as a function of sample time, τ , for $\nu_{\text{beat}} = 3165$ MHz.

TABLE II

τ (seconds)	1	2	5	10	25
m	101	51	26	26	11

Sample size, m, as a function of sample time, τ , for $\nu_{\text{beat}} = 593$ MHz.

TABLE III

τ (seconds)	0.1	0.2	0.4	0.7	1.0	2	5	10	25	50	100	250	500	1000
m	101	101	101	101	101	51	26	26	11	3	3	3	3	2

Sample Size, m, as a Function of Sample Time, τ , for $\nu_{\text{beat}} = 2698$ MHz

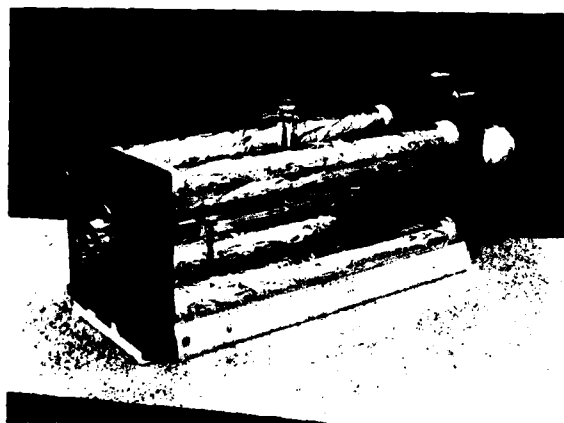
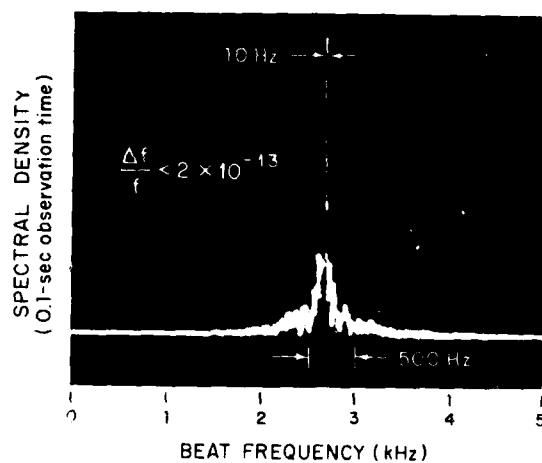
TABLE IV

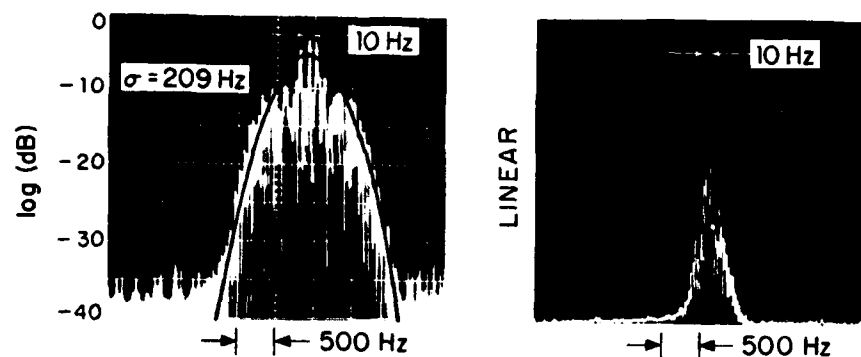
VIBRATIONAL - ROTATIONAL CONSTANTS OF CO₂ ISOTOPE LASER TRANSITIONS

Isotope Constants in MHz	¹² C ¹⁶ O ₂	¹³ C ¹⁶ O ₂	¹³ C ¹⁸ O ₂	¹⁴ C ¹⁶ O ₂
V (0)001-I	28 808 813.8	27 383 792.4	27 838 551.2	25 965 920.7
V (0)001-II	31 889 960.2	30 508 659.6	30 785 884.8	29 460 002.6
P001	11 606.2072	11 610.1771	10 319.0962	11 613.6722
B _I	11 697.5697	11 683.4536	10 403.4743	11 674.7428
B _{II}	11 706.3645	11 719.3785	10 398.9836	11 727.0591
D ₀₀₁	3.9883 × 10 ⁻⁵	3.9923 × 10 ⁻⁵	3.1451 × 10 ⁻⁵	3.9879 × 10 ⁻⁵
D _I	3.4462 × 10 ⁻⁵	3.6126 × 10 ⁻⁵	2.7147 × 10 ⁻⁵	3.8183 × 10 ⁻⁵
D _{II}	4.7115 × 10 ⁻⁵	4.7567 × 10 ⁻⁵	3.6544 × 10 ⁻⁵	4.8317 × 10 ⁻⁵
H ₀₀₁	5.15 × 10 ⁻¹⁰	1.012 × 10 ⁻⁹	-8.12 × 10 ⁻¹⁰	2.852 × 10 ⁻⁹
H _I	5.730 × 10 ⁻⁹	7.159 × 10 ⁻⁹	1.807 × 10 ⁻⁹	3.7293 × 10 ⁻⁸
H _{II}	6.984 × 10 ⁻⁹	9.022 × 10 ⁻⁹	4.018 × 10 ⁻⁹	1.6284 × 10 ⁻⁸

$$c = 299\,792\,458 \text{ m/sec}$$

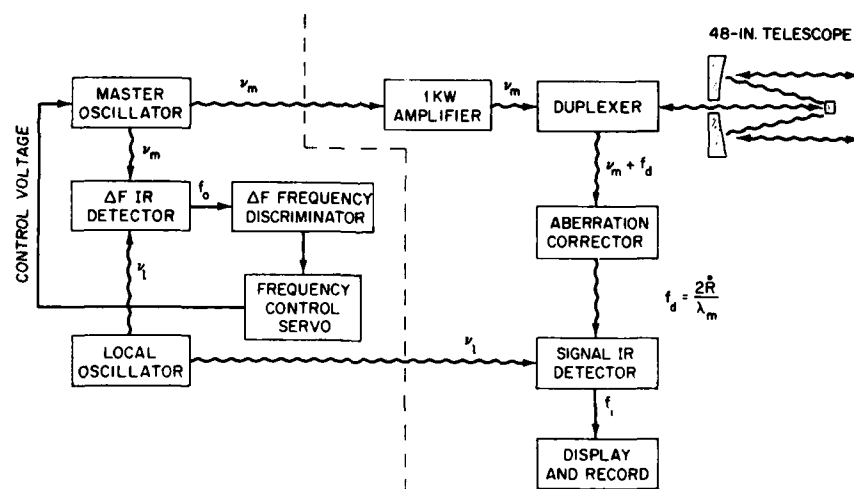
TABLE V


FIG. 1 0.5 meter, internal mirror, piezoelectrically tuned stable CO₂ laser.

FIG. 2. Real-time spectral density of the beat note of two free running, 0.5 meter stable CO₂ lasers for 0.1 second observation time; horizontal scale: 500 Hz/cm; resolution 10 Hz.



BEAT NOTE SPECTRA OF TWO CO₂ LASERS

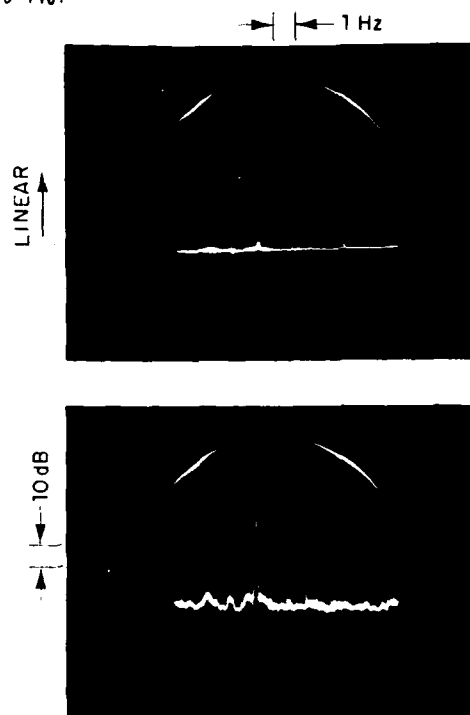
FIG. 3 A beat note similar to Figure 2 but measured under somewhat noisier environmental conditions.



INFRARED RADAR SYSTEM BLOCK DIAGRAM

FIG.4 10.6μm radar system block diagram.

-5-7481



RF SPECTRUM OF CO₂ LASER BEAT
WITH FREQUENCY SERVO
(0.02 Hz Resolution)

FIG. 5 10 MHz beat note of two offset-locked lasers;
spectrum analyzer resolution set to 0.02 Hz.

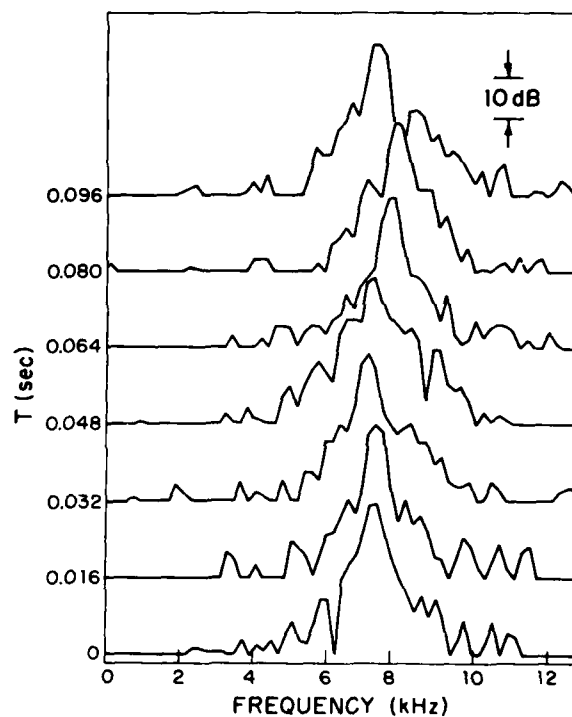
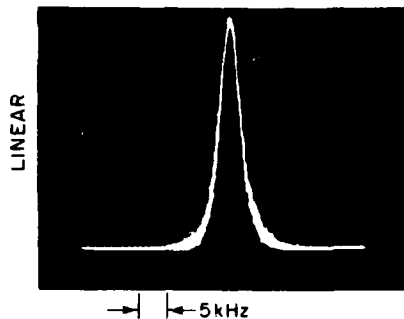


FIG. 6 10.6 μ m radar return spectra from GEOS III
geodetic satellite.



Fig. 7. 1.5 meter, internal-grating-controlled, piezoelectrically tuned
stable CO₂ laser.



3,165 MHz BEAT FREQUENCY OF A
 $^{12}\text{C}^{16}\text{O}_2$ 001-I BAND R(18) TRANSITION
 WITH A $^{14}\text{C}^{16}\text{O}_2$ LASER LINE

I. F. BANDWIDTH : 3 kHz

Fig. 8. Spectrum analyzer display of 3,165 MHz beat frequency of the $^{14}\text{C}^{16}\text{O}_2$ 0001 - $[10^00, 02^00]_{II}$ band P(10) and the $^{12}\text{C}^{16}\text{O}_2$ 0001 - $[10^00, 02^00]_{I}$ band R(18) transitions; I. F. Bandwidth: 3 kHz.

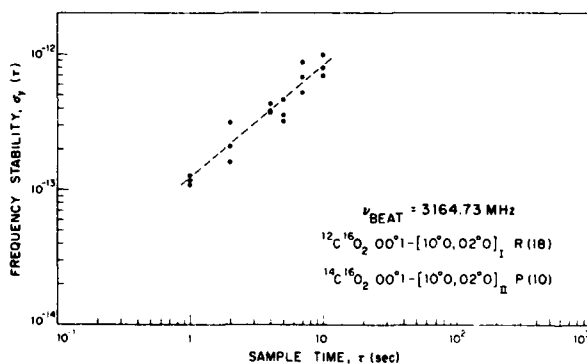


FIG. 9 Time domain stability of the beat note shown in Figure 8.

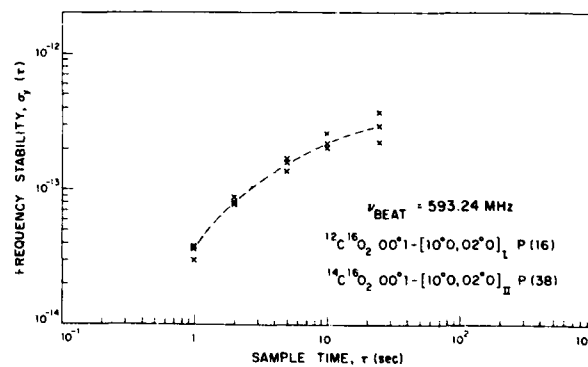


FIG. 10 Time domain stability of the 593.24 MHz self-beat in a $^{14}\text{C}^{16}\text{O}_2$ - $^{12}\text{C}^{16}\text{O}_2$ combination laser.

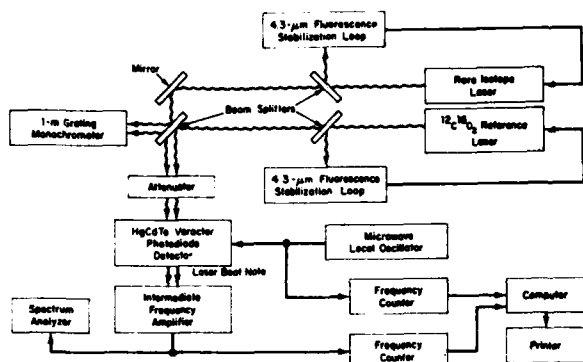


FIG. 11 Two-channel heterodyne system used to determine CO_2 isotope laser frequencies.

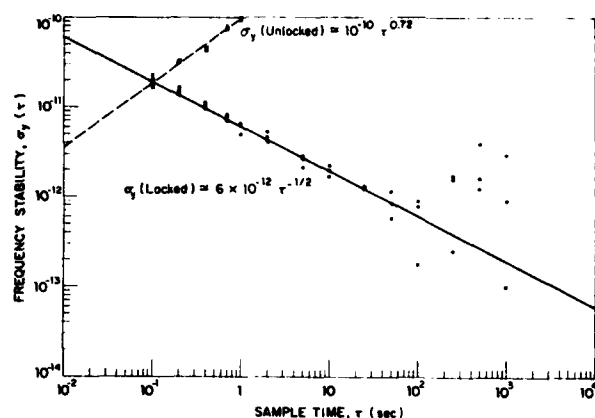


FIG. 12 Time domain stability of independently locked grating controlled CO_2 lasers.

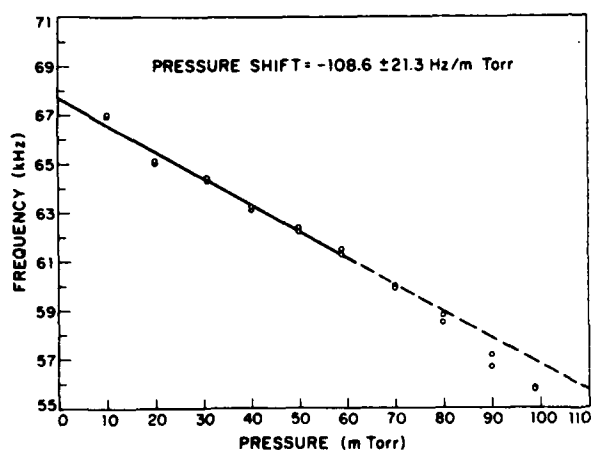


FIG. 13 Variation of the $2,697,862 \pm 6$ kHz beat frequency as a function of pressure change in the $^{12}\text{C}^{16}\text{O}_2$ reference cell for the $00^01 - [10^00, 02^00]_1$ band P (20) transition.

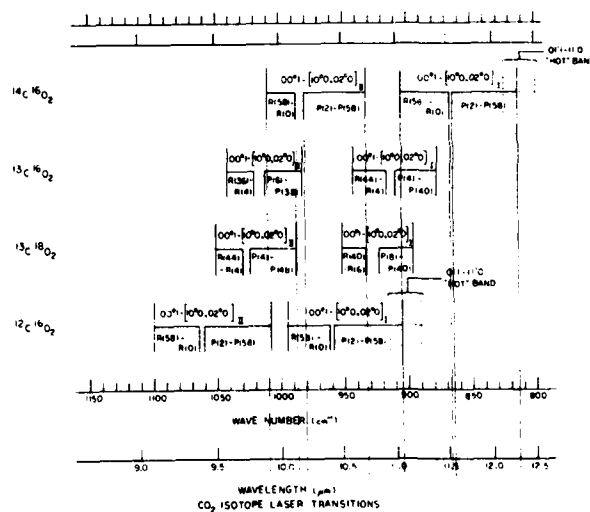


FIG. 14 Comparison of the frequency and wavelength domains of rare CO_2 isotope lasers with $^{12}\text{C}^{16}\text{O}_2$.

ELECTRONIC TUNING AND PHASE-LOCK TECHNIQUES FOR OPTICALLY PUMPED FAR INFRARED LASERS

S. R. Stein
Frequency and Time Standards Section
National Bureau of Standards
Boulder, Colorado 80302
(303) 499-1000, extension 3224

and

H. van de Stadt*
Joint Institute for Laboratory Astrophysics
National Bureau of Standards and University of Colorado
Boulder, Colorado 80309

Summary

An electronic tuning technique for far-infrared (FIR) lasers has been developed which makes it possible both to frequency modulate a laser at a very high rate and to phase-lock it to a reference oscillator. Tuning is achieved by applying an electronic (Stark) field to the molecules in an optically pumped FIR laser. Since such lasers operate on rotational-vibrational transitions of molecules, they exhibit either first- or second-order Stark effects which result in an effective broadening of the molecular transition. This introduces a change in the cavity pulling, which is present in any laser in which the cavity resonant frequency is offset from the molecular transition.

Using a CH_3F laser at a frequency of 604 GHz (496 μm), we have observed (by direct measurements with a metal-insulator-metal diode) a maximum static electronic tuning of 500 kHz with a maximum sensitivity of 5 kHz/(V/cm). The tuning rate should be limited only by the lifetime of the photons in the laser cavity. We have observed frequency modulation rates of 300 kHz and a modulation index greater than one has been obtained at a modulation rate of 50 kHz.

We report the ability to phase lock such a tunable FIR laser to a reference laser at an offset frequency of the order of 500 kHz. This was achieved by using the Stark-tuning method to correct for fast frequency fluctuations, up to approximately 20 kHz, and a piezo-electric tuner (PZT) to slowly correct for changes in the laser cavity length.

The tuning and phase-lock methods reported here are needed for several important applications, for example: infrared frequency synthesis, generation of time from infrared (and higher) frequency standards, directional and synthesized infrared communications, and FM LIDAR.

Experimental Description

A schematic diagram of the experimental setup is shown in Figure 1. A CO_2 laser and amplifying tube provide the pump beams, at a wavelength of approximately 10 μm , necessary to excite a particular vibrational level of the molecules in FIR lasers I and II. One laser, FIR-I, serves as a reference laser. A metal-insulator-metal diode with a 50 μm diameter tungsten whisker and a polished nickel post is used for heterodyning the two FIR frequencies. More details about the setup can be found in Ref. 1.

*Visiting Fellow 1976-77, on leave from Astronomical Institute, Utrecht, The Netherlands.

The FIR-II laser consists of a glass tube, 1.04 m long and 32 mm internal diameter. Two lengths of copper tape are attached over the total length of the tube to serve as the internal Stark electrodes. Both ends of the tube are closed by flat, metal mirrors, of which one is movable in two different ways: A PZT is used for fine movements (4 μm maximum), while a micrometer drive provides the coarse motion.

A block diagram of the phase-lock loop system is shown in Figure 2. The two FIR lasers produce a beat frequency on the order of several hundred kilohertz in the metal-insulator-metal diode detector. After amplification and filtering this beat note is fed into a double-balanced mixer, where it is compared to the frequency of a synthesizer. The phase-lock loop tunes the FIR-II laser so that the heterodyne signal between the two lasers and the reference signal from the synthesizer are in approximate phase quadrature. This is accomplished in two different ways: Fast tuning is achieved by amplifying and filtering the detected phase-difference signal and then feeding it to the Stark electrodes. The attack time in this feed-back loop is approximately 10 μs . Slow tuning, over a larger range, is mainly necessary to compensate for length changes of the FIR-II laser cavity due to temperature changes. This tuning is achieved by feeding the detected phase-difference signal through an integrating amplifier and applying it to the PZT-tuning element of the FIR-II laser. Unwanted interactions between the two feedback loops are not present because of the large difference in speed (10^4) of the two mechanisms. The present system stays in a phase-locked condition for about 15 minutes. This time is mainly limited by the frequency instability of the CO_2 -pump laser and the mechanical instability of the diode, rather than by the dynamic range of the phase-lock loop itself.

Measurements

All measurements are preceded by mechanically tuning the FIR-II laser to a dominant longitudinal mode. In Figure 3, we show the output power variation as a function of cavity length for the Stark tunable laser. The two main maxima occur at a cavity length difference of 248 μm , half the wavelength of the 496 μm transition in CH_3F . The relatively clean mode spectrum of this laser avoids the otherwise annoying troubles due to mode interference.

The next step is to first detune the cavity slightly from the transition line center and then apply a dc voltage of about 400 V to the Stark electrodes. This is done while the FIR-I laser remains tuned to its line center, allowing one to monitor the frequency changes of FIR-II by direct heterodyning in a frequency region of the order of 500 kHz. The initial cavity detuning produces a frequency shift of

300 to 600 kHz in the FIR-II laser, while the additional amount of electronic tuning, which is proportional to the initial cavity detuning, has a maximum sensitivity of 5 kHz/(V/cm). Associated with the frequency change there may be an amplitude change with a relative magnitude of 0.01/(V/cm).¹

The application of an alternating field, superimposed on the dc electric field, produces both amplitude and frequency modulation sidebands. The ratio of AM sidebands to the first FM sidebands is expected to be $2 \times 10^{-6} f_m$, where f_m is the modulation frequency in Hz, as calculated from the sensitivities quoted in the preceding paragraph.¹ It has been verified qualitatively that the AM and FM sidebands have indeed equal strength at $f_m = 500$ kHz.

Figure 4 shows a typical frequency spectrum of a frequency-modulated, Stark-tuned FIR laser. It was taken using a spectrum analyzer connected to the amplified heterodyne signal from the diode. Since only the FIR-II laser is modulated, it represents the power density of this laser as a function of frequency. This particular spectrum was taken while the laser was modulated at 50 kHz. The relative magnitude of the sidebands is typical for frequency modulation with a deviation (half the peak-to-peak frequency shift) equal to 75 kHz. The AM sidebands in this case are ten times less powerful than the FM sidebands, in agreement with our earlier statement and indicating that we have a nearly pure frequency modulated spectrum.

Observation of the beat note between the two FIR lasers indicates that the 3 dB line width of the unlocked lasers is a few kHz. A typical spectrum is shown in Figure 5a. In order to phase lock such oscillators it is necessary that all the amplifiers and filters in the loop be broadband compared to the free-running line width. When the phase-lock loop is operating, the line width of the beat note decreases dramatically as can be seen in Figure 5b. If a stable reference were substituted for the FIR-I laser, then this change would represent a proportional increase in stability of the tunable laser.

The most important parameter describing the phase-lock loop is its attack time, i.e., the inverse of the unity gain angular frequency. It can be measured by phase modulating the synthesizer which generates the frequency offset between the two lasers. When the frequency of this phase modulation is small compared to the unity gain frequency, coherent modulation sidebands of constant amplitude are impressed on the tunable laser. However, when the modulation frequency is large compared to the unity gain frequency, the tunable laser can't follow the phase modulation of the reference. We observe this phenomenon in both the spectrum of the beat between the two lasers and the spectrum of the control voltage to the tunable laser. Our best estimate of the unity gain frequency is 18 kHz which corresponds to an attack time of approximately 10 μ s.

In order to analyze the operation of the servo system it is useful to examine the spectrum of the feedback voltage to the Stark tuning electrodes. An example of such a spectrum is shown in Figure 6. The primary value of this measurement is to detect instabilities in the loop which show up as peaks in the spectrum. The loop is observed to be stable, but begins to oscillate when the gain is increased by 10 dB. It is also possible to use the spectrum of the feedback voltage to determine $S_\phi(f)$ for the pair of oscillators. This is discussed in more detail in the next section.

When the Stark tuner is operating alone, the dynamic range of the loop is very limited. For example,

a frequency change of the reference oscillator in excess of 50 kHz causes the loop to come out of lock. This dynamic range is insufficient to control the slow frequency changes in the FIR lasers which we surmise are due to temperature changes. The inclusion of the PZT tuner, which can move one end mirror ± 2 μ m, results in an order of magnitude improvement in the dynamic range.

Interpretation

We selected the 496 μ m (604 GHz) laser transition in CH₃F for this study because it is a relatively strong line and the Stark effect in CH₃F is linear and well understood. The CO₂ laser selectively pumps the molecules first to an excited vibrational level and the resulting population inversion induces the 496 μ m rotational transition to lase. An electric field splits the rotational transition into several components, each with a known frequency shift proportional to the field strength. From the known line width of the individual components one can show^{1,2} that the net effect of the Stark field is an overall broadening of the line of approximately 30 kHz/(V/cm).

The tuning of the laser frequency is produced by this broadening via the effect of "cavity pulling."³ The actual laser frequency, ν , is determined by:

$$\frac{\nu - \nu_m}{\Delta \nu_m} = \frac{\nu - \nu_c}{\Delta \nu_c}$$

where ν_c is the resonant frequency of the empty laser cavity and $\Delta \nu_c$ its full width, ν_m is the frequency of the molecular transition and $\Delta \nu_m$ its full width. It is clear from this equation that an offset between ν_c and ν_m produces a static frequency shift of the laser frequency towards the cavity line center. The amount of shift depends on the cavity offset itself, as well as the magnitudes of $\Delta \nu_c$ and $\Delta \nu_m$. The very last dependence is used in our experiments.

Differentiation of the pulling equation with respect to the applied field E yields:

$$\frac{d\nu}{dE} = \frac{\Delta \nu_c (\nu_c - \nu_m)}{(\Delta \nu_m + \Delta \nu_c)^2} \frac{d}{dE} (\Delta \nu_m)$$

The shift is thus proportional to the initial offset of the cavity from line center and the proportionality factor can be calculated from known quantities. We find in our case a maximum value of 5 kHz/(V/cm), in good agreement with measured values.

The operation of the phase-lock loop, described earlier, is best understood in terms of the open-loop transfer function $G_{eq}(j\omega)$.⁴ Let us call the open-loop phase fluctuations of the two lasers ϕ_I and ϕ_{II} , respectively. Then the closed-loop phase fluctuations between the two lasers, $\Delta \phi$, is

$$\Delta \phi = \frac{\phi_{II} - \phi_I}{1 + G_{eq}}$$

where the frequency synthesizer is assumed to have negligible phase fluctuations. The open-loop transfer function is the product of the tuning rate of the FIR-II laser, the sensitivities of the MIM diode and the double balanced mixer, the gain of amplifiers G_1 and G_2 , and the low pass transfer functions F_1 , F_2 and F_3 . The asymptotic form is shown in Figure 7. The break point at f_1 is due to the low pass filter (F_2) which decreases the feedback of the second harmonic of the offset frequency

between the two lasers. The break point at f_3 is due to the finite band width (F_3) of the amplifiers. The phase shift caused by these two low pass filters currently limits the maximum unity gain frequency to about 20 kHz. The break at f_4 is due to the PZT tuner and has negligible influence on the second-to-second performance of the loop.

The voltage, V , measured at the spectrum analyzer is easily related to the open-loop phase fluctuations between the two lasers. Since the spectrum analyzer follows filter F_2 ,

$$V = \frac{K(\phi_{II} - \phi_I)}{(1+jf/f_2)(1+G_{eq})}$$

where K is a constant. Solution of this equation indicates that the slope of the voltage spectrum should change by f^2 at a frequency near the unity gain frequency. The break does not occur exactly at the unity gain frequency mainly due to the fact that the poles at f_2 and f_3 are not far removed. Examination of the spectrum in Figure 6 shows just such a slope change at 18 kHz. It also shows a slope change at 4 kHz, below which the voltage spectrum is constant. This behavior is consistent with the interpretation of 4 kHz as the fast linewidth (approximately equal to the 3 dB linewidth) of the unstabilized lasers.⁵

Since the time constants of the various filters in this phase-lock loop are not widely spaced, $S_\phi(f)$, the spectral density of phase fluctuations between the two lasers, must be calculated on a point-by-point basis. We find that $S_\phi(30 \text{ kHz}) \approx -90 \text{ dB}$ relative to 1 rad^2 . This noise level is comparable to a Gunn-effect oscillator referred to its operating frequency of 10 GHz.

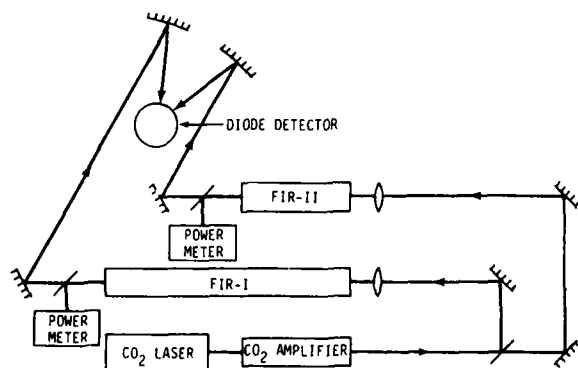


FIGURE 1. Schematic diagram of the system for heterodyning two far infrared (FIR) lasers.

Conclusions

In the near future, we expect the expansion of the fields of time and frequency and telecommunications into the infrared to accelerate significantly. It is likely that the far-infrared laser will be a workhorse oscillator for that frequency range. To promote this possibility, it is necessary to develop a variety of modulation and phase-locking techniques such as are available in the microwave region. The Stark frequency modulation and phase-lock technique which is described here is a step in this direction. When combined with suitable low noise microwave sources and efficient multipliers it will permit frequency synthesis over a large portion of the far-infrared wavelength region.

References

1. S. R. Stein, Allan S. Risley, H. Van de Stadt and F. Strumia, to be published in *Applied Optics*, July 1977.
2. D. T. Hodges, J. R. Tucker and T. S. Hartwick, *Infrared Physics* **16**, 175 (1976).
3. J. Viennet, C. Audoin, M. Desaintfuscien, *Proc. 25th Annual Symposium on Frequency Control (Elec. Ind. Assoc., Washington, DC, 1971)*, p. 337.
4. James L. Melsa and Donald G. Schultz, *Linear Control Systems* (McGraw-Hill, New York, 1969), pp. 94-119.
5. Donald Halford, *Proc. of the Frequency Standards and Metrology Seminar (Universite Laval, Quebec, Canada, 1971)*, pp. 442-446.

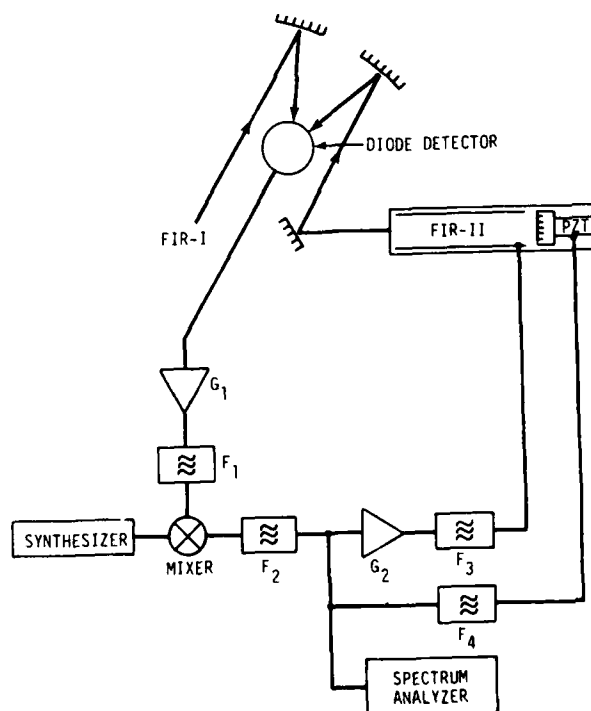


FIGURE 2. Block diagram of the phase-lock loop system.

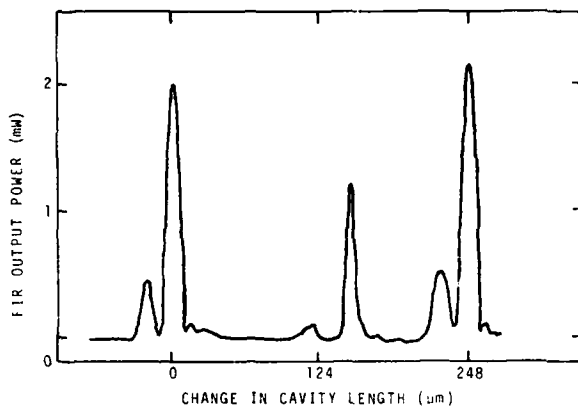


FIGURE 3. Output power of the FIR laser as a function of cavity length. In addition to the main cavity modes, there are visible a small number of higher order modes.

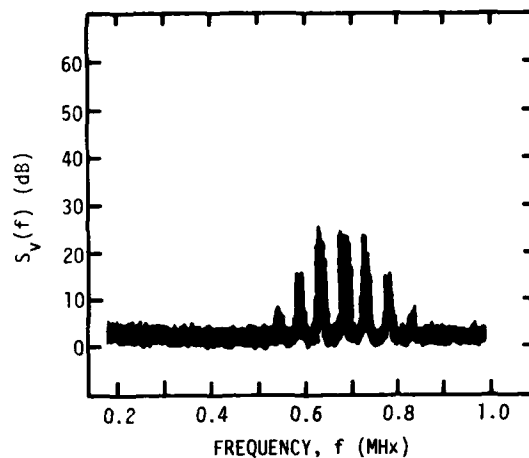


FIGURE 4. Frequency spectrum of a Stark-tuned FIR laser with a modulation frequency of 50 kHz.

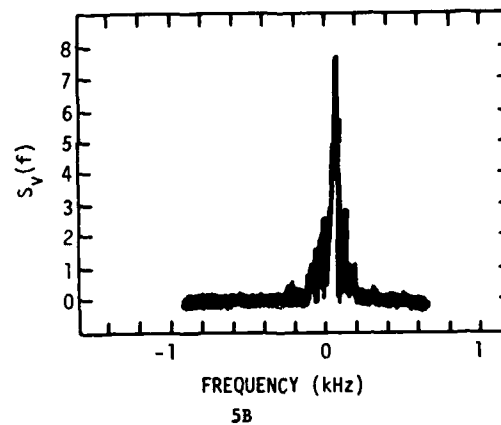
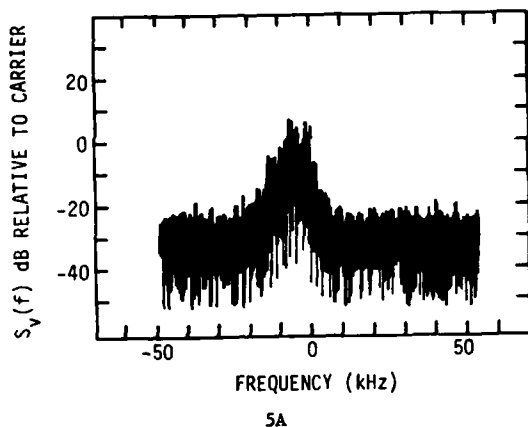


Fig. 5. a) Beat frequency spectrum of two unlocked FIR lasers.
b) Beat frequency spectrum of two phase-locked FIR lasers.

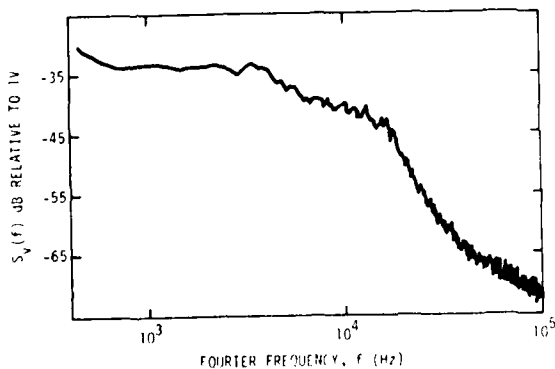


FIGURE 6. Frequency spectrum of the feedback voltage in the closed-loop condition.

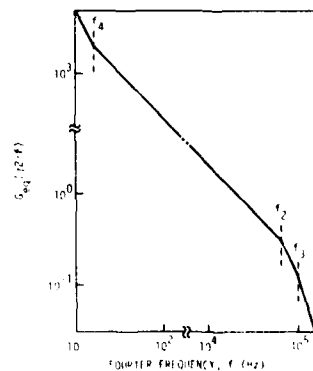


FIGURE 7. Asymptotic form of the magnitude of the open-loop transfer function.

FREQUENCY MODULATION OF A FAR INFRARED
 CH_3F LASER BY STARK EFFECT*

by

R. Benedetti, A. Di Lieto, M. Inguscio
P. Minguzzi, F. Strumia, M. Tonelli

Istituto di Fisica dell'Università - Pisa
Pisa, Italy

Summary

We report a direct frequency measurement of the shifts and modulation induced by the application of an electric field on an optically pumped CH_3F far infrared laser. These effects are obtained only if the laser resonator is slightly offset from the molecular transition frequency, so that the electric field enhances the cavity pulling shift. A detailed computation of the Stark multiplet pattern and intensities is included and the change in the gain profile of the active medium by the electric field is discussed. The demonstrated control of the laser frequency by an electric field will allow high speed FM modulation or fast laser frequency stabilization.

Introduction

Many new far infrared laser lines have been obtained by optically pumping near infrared vibrational transitions of several molecules. The frequency of these lasers ranges from 0.15 to about 20 THz and a CW intensity from 1 to 100 mw can be easily obtained for the strongest lines. The frequency tunability is within the Doppler or collision width of the lasing line and also the speed of the frequency modulation is limited because of the absence of suitable electro-optic materials in the far infrared.

The Stark effect was proposed^{1,2} as a possible method to increase the frequency tunability and to obtain a high speed frequency modulation in FIR lasers, since many of the known lasing molecules exhibit a large Stark effect. Laser action in the presence of an electric field has been previously reported in several papers.

At present at least two groups^{10,11}, to our knowledge, are at work about frequency measurements on Stark controlled lasers, however no report has yet appeared. Here we report a direct frequency measurement of Stark tunability on the 604 GHz line of CH_3F laser.

Experimental apparatus

In fig. 1 the experimental set-up is

shown schematically. It comprises: a CO_2 laser with the frequency stabilization servo for pumping, a CH_3F Stark controlled laser, a 75 GHz klystron with the frequency measurement chain, a W-Ni point contact mixer and a spectrum analyzer to observe the beat note. Most of the details of this apparatus have been described in previous papers^{10,11} so only the main features will be discussed here.

The FIR laser has a dielectric waveguide configuration with flat internal mirrors and coupling holes for input and output power. As a waveguide we use a pyrex tube, 30 cm long and 13 mm internal diameter. The Stark plates are two silver coated glass plates, placed at a distance of 1 cm outside of the vacuum system. They are slightly shorter than the waveguide, so small portions of the active medium, near input and output holes are not immersed into the electric field. The CH_3F gas is flowed slowly through the laser at a pressure of about 50 mTorr. The pump power is delivered by a commercial CO_2 laser, capable of about 20 watt at 9.55 μm , and is focused into the FIR resonator by concave mirror optics. A small fraction of the CO_2 power is passed through a CH_3F reference cell and is monitored by a pyroelectric detector. The signal from this detector is used to generate an error voltage that locks the CO_2 frequency to the absorption peak of CH_3F^{13} . In most of our measurements the polarization of the CO_2 laser is parallel to the Stark field. The FIR laser output is mixed in a W-Ni open structure point contact diode with the microwave power from a klystron. The klystron is operated free running at about 75530 MHz and is energized by a very low noise, low drift power supply. Part of the microwave power, via a directional coupler, is used to compare the klystron frequency to a 5 MHz crystal oscillator by means of a synthesizer and a conventional frequency multiplier chain. The beat note between the laser and the eighth harmonics of the klystron at about 57.5 MHz is processed by a wideband amplifier and displayed on a spectrum analyzer.

Results and Discussion

The first result we obtained is a new

direct measurement of the $496 \mu\text{m}$ $K = 2$ line of the CH_3F laser. The experimental apparatus was slightly different with the klystron phase-locked to a stable reference obtained by multiplication from a very high quality quartz oscillator. The details of this measurement are reported in ref 12 and the result is

$$f_{\text{CH}_3\text{F}}(496 \mu\text{m}, K=2) = 604297.5 \pm 0.2 \text{ MHz}$$

Occasionally we observed on the spectrum analyzer a double peak in the beat signal; a typical result is shown in fig. 2.

The two peaks are of equal intensity, are separated of about 920 KHz and both intensity and frequency separation seem to be unaffected by pressure changing in laser operation (from 40 to 150 mTorr) and by a change in the frequency offset between the CO_2 pump radiation and the CH_3F absorption transition.

The two-peak situation has been observed both in copper and in dielectric waveguide resonators. The two peaks are too symmetric and their frequency separation too small to be interpreted as different modes of the laser cavity.

It could be guessed an interpretation according to the theory developed by De Temple and Danielewicz¹⁴. Kramer and Weiss¹⁵ have also reported something similar. They write of a central dip 1 MHz wide. Considering that their laser threshold is lower than ours, the two peaks we observe could be the upper structure of their dip. Nevertheless the frequency separation we observe between the peaks is too large for the offset between the CO_2 radiation and the CH_3F absorption transition, which is of 11~13 MHz. Another interpretation could be guessed by the theory reported by Leite et al. in ref. 16.

In our experiment the pump transition is "Doppler broadened" but the Doppler effect ($\Delta\nu_D \approx 1.3 \text{ MHz}$) at the FIR frequency is negligible compared to the pressure broadening ($\Delta\nu_p \approx 2 \text{ MHz}$). In this case no splitting of the emitted FIR line is predicted¹⁶. It is worth noting that the pressure broadening of the ν_3 , $K=2$, $J=12 \rightarrow J=11$ rotational transition in CH_3F is extrapolated here (and also in 14) by measurements of pressure broadenings in the ground vibrational state at low J values¹⁷. So the value which is given (40 MHz/Torr) could be significantly overestimated.

The experimental results on the Stark control of the laser frequency can be better understood if some of the theoretical results of the next section are anticipated. The gain curve of the active medium is symmetrically broadened but not appreciably shifted

by a moderate electric field, so if the laser resonator is tuned to the center of the molecular transition, no frequency change will be induced by the electric field.

So a certain amount of cavity pulling is initially introduced by offsetting the frequency of the laser cavity from the molecular transition. The FIR frequency shift or modulation is then obtained by applying the electric field to the Stark plates. A typical result for a static frequency shift is shown in fig. 3. In that case the applied field is 220V/cm and the observed frequency shift (upper trace) is about 200 KHz. The vertical scale is logarithmic and the signal-to-noise ratio is about 10 dB. The maximum shift we observed is about 300 KHz with a Stark coefficient (frequency shift/applied electric field) of 2 KHz/(volt/cm).

The observed shift is roughly proportional to the cavity detuning and to $|E|$ and its direction is the same as the cavity detuning itself.

If an AC electric field is superimposed to the DC field a frequency modulation of the laser output is obtained. However, due to the free running operation of the klystron, a poor resolution on the spectrum analyzer is available (see fig. 3); so at present only qualitative results have been obtained.

Theory

The CH_3F molecule is a symmetric top and simple equations can be obtained for the frequencies of the rotational and vibrational lines in presence of an electric field \vec{E} (Stark effect). Moreover for CH_3F the molecular constants have been accurately measured¹⁸ and accurate computation of the Stark effect can be performed. The selection rules for the rotational transitions are $\Delta J = \pm 1$, $\Delta K = 0$, and $\Delta M = 0, \pm 1$ and the frequency shifts of the Stark components are:

$$\Delta\nu_R = \frac{2\mu_e EMK}{J(J^2-1)h} - \frac{\mu_e K \Delta M E}{J(J+1)h} \quad (1)$$

where J is the larger of the two quantum numbers involved.

Terms in E^2 are negligible for CH_3F up to a field of several KV/cm and for $\nu_3 = 1$, $J=12-11$ the eq.(1) becomes:

$$\Delta\nu_R = KE(1.118M - 6.149 \Delta M) \frac{\text{KHz}}{\text{V/cm}} \quad (2)$$

For the vibrational transitions of the Q branch ($\Delta J = 0$) we found:

$$\Delta\nu_V = \frac{(\mu_o - \mu_e) E K M}{J(J+1)h} - \frac{\mu_e E K \Delta M}{J(J+1)h} \quad (3)$$

and for the Q(12) line of CH_3F

$$\Delta \nu = K(-0.151 M - 6.150 \Delta M) E \frac{\text{KHz}}{\text{V/cm}} \quad (4)$$

where μ_0 is the electric dipole moment of the ground vibrational state and μ_e is the electric dipole moment in the excited state. The shifts of the vibrational lines are much smaller than the line widths as a consequence of the term $(\mu_0 - \mu_e)$ and the pumping efficiency of the CO_2 laser radiation is not significantly changed in presence of an electric field.

The position of the rotational lines as a function of the electric field is shown in fig. 4. The electric dipole moment in the excited vibrational state and the \vec{E} vector of the FIR radiation are orthogonal to the \vec{E} vector of the CO_2 radiation as a consequence of the pumping by a Q line. When the pump is obtained by a $\Delta M = 0$ transition ($\vec{E}_{\text{CO}_2} // \vec{E}_{\text{Stark}}$) the FIR radiation is emitted by $\Delta M = \pm 1$ Stark transitions ($\vec{E}_{\text{FIR}} \perp \vec{E}_{\text{Stark}}$). The relative intensity of the Stark components of the rotational $J \rightarrow J + 1$ line are given by the equations

$$I(\Delta M = 0) = (J + 1)^2 - M^2 \quad (5)$$

$$I(\Delta M = \pm 1) = (J + 1 \pm M)(J + 2 \pm M) \quad (6)$$

For $\Delta M = 0$ we obtain 23 lines, for $\Delta M = \pm 1$ we obtain 46 lines but we can see from eq. 1 that 12 lines with $\Delta M = -1$ (from $M = -11$ to $M = 0$) have the same frequency as 12 lines with $\Delta M = +1$ (from $M = 0$ to $M = +11$) and there are only 34 resolved lines of which only the 12 central lines have a significant intensity as shown in fig. 5.

The gain profile of the FIR laser can be computed from the above equations if the shape and the width of the $496 \mu\text{m}$ lines are known. At room temperature the Doppler width is 1.3 MHz. No experimental measurements exist for the pressure self-broadening but from the discussion of preceding section it seems reasonable to assume a pressure broadening of $\approx 40 \text{ MHz/Torr}$. At the best working pressure we have an inhomogeneous pressure-broadened line which is a little larger than the Doppler width. The shape of the line is given by the Voigt equation but as a good approximation we can use a Lorentz's line with a width $2\gamma \approx 2 \text{ MHz}$. In this case, assuming an equal excitation for all the M sublevels, the gain profile of the FIR laser is given by

$$I_{\Delta M=0}(\nu) = \frac{\gamma^2}{\sum_M I_{M_0}} \sum_M \frac{I_{M_0}}{(\nu - \nu_M)^2 + \gamma^2}$$

and

$$I_{\Delta M=\pm 1}(\nu) = \frac{\gamma^2}{\sum_M I_{M\pm}} \sum_M \left(\frac{I_{M+}}{(\nu - \nu_{M+})^2 + \gamma^2} + \frac{I_{M-}}{(\nu - \nu_{M-})^2 + \gamma^2} \right)$$

The results for different values of the applied electric field are shown in fig. 6. We can see that the single Stark components start to be resolved for an electric field of about 1 KV/cm. Moreover the gain for a given E field is a little larger for the $\Delta M = \pm 1$ transitions.

The threshold of our present dielectric waveguide laser is at about 0.3 of the gain profile and, as a consequence, the Stark effect tunability of the FIR laser can be obtained only via frequency pulling of the offset resonator. New FIR lasers are in preparation which should lower the threshold by a factor 4 ~ 5. In this case a Stark tunability of several MHz could be easily accomplished.

The present difficulty follows from the high J value of the lasing transition which reduces the relative intensity of the single Stark components. The Stark shift is also strongly dependent on the J and K values. A low J and high K transition will have a much larger Stark shift and will be easier to lase on a single Stark component. In this case a tunability range of several hundreds of MHz will be feasible.

Acknowledgement

We want to express our appreciation to Mr. M. Montanari of the mechanical shop and Mr. M. Francesconi of the electronic shop for their enthusiastic and skilful collaboration. We also thank "Istituto Elettrotecnico Nazionale" of Torino for the loan of the CO_2 laser.

References

- *Work supported by Gruppo Nazionale Struttura della Materia del CNR, Pisa, Italy
1. M. Inguscio, P. Minguzzi and F. Strumia, Conf. Dig. Int. Conf. Infrared Phys., Zurich, Aug. 1975
2. M. Inguscio, P. Minguzzi and F. Strumia, Infrared Phys. 16 (1976) 453
3. H.R. Fetterman, H.R. Schlossberg and C.D. Parker, Appl. Phys. Lett. 23 (1973) 684
4. F. Strumia and P. Minguzzi, Proc. 2nd Symposium on Freq. Standards and Metrology, Copper Mountain (1976) p. 65

5. F. Strumia, R. Benedetti, M. Inguscio, P. Minguzzi, M. Tonelli, E. Bava, A. De Marchi, A. Godone, Conf. Dig. 2nd Int. Submillimeter Waves Conf., Puerto Rico, Dec. 1976, p. S-9
6. H.R. Fetterman, C.D. Parker and P.E. Tannenwald, Opt. Commun. 18 (1976) 10
7. M.S. Tobin and R.E. Jensen, Appl. Opt. 15 (1976) 2023
8. M.S. Tobin and R.E. Jensen, Conf. Dig. 2nd Int. Submillimeter Waves Conf., Puerto Rico, Dec. 1976, p. 167
9. K. P. Koo and P.C. Claspy, Conf. Dig. 2nd Int. Submillimeter Waves Conf., Puerto Rico, Dec. 1976, p. 171
10. M. Inguscio, P. Minguzzi and M. Tonelli, Opt. Commun., to be published.
11. S.R. Stein, A.S. Risley, H. Van de Stadt and F. Strumia, Appl. Opt., to be published.
12. E. Bava, A. De Marchi, A. Godone, R. Benedetti, M. Inguscio, P. Minguzzi, F. Strumia and M. Tonelli, Opt. Commun. 21 (1977) 46.
13. P. Minguzzi and M. Tonelli, J. Phys. E : Sci. Instr., to be published
14. T.A. De Temple and E.T. Danielewicz IEEE J. Quant. El. QE - 12 (1976) 40
15. G. Kramer and C.O. Weiss, Appl. Phys. 10 (1976) 187
16. J.R.R. Leite, D. Seligson, J.J. Mickey, M. Ducloy, A. Sanchez and M.S. Feld, Conf. Dig. 2nd Int. Submillimeter Waves Conf., Puerto Rico, Dec. 1976, p. 83
17. G. Birnbaum, E.R. Cohen, J.R. Rusk, J. Chem. Phys. 49 (1968) 5150
18. S.M. Freund, G. Duxbury, M. Romheld, J.T. Tiedje and T. Oka, J. Mol. Spectr. 52 (1974) 38

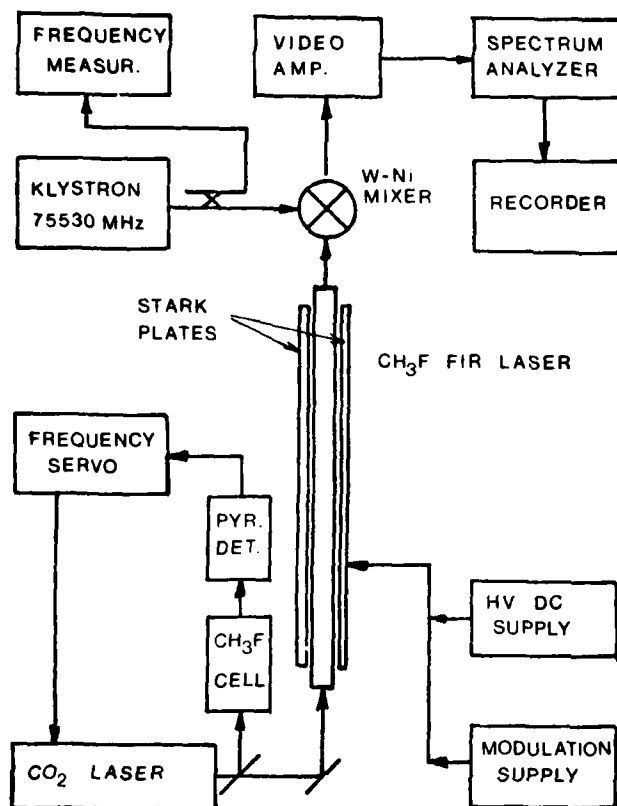


Fig. 1 Schematic diagram of the experimental apparatus

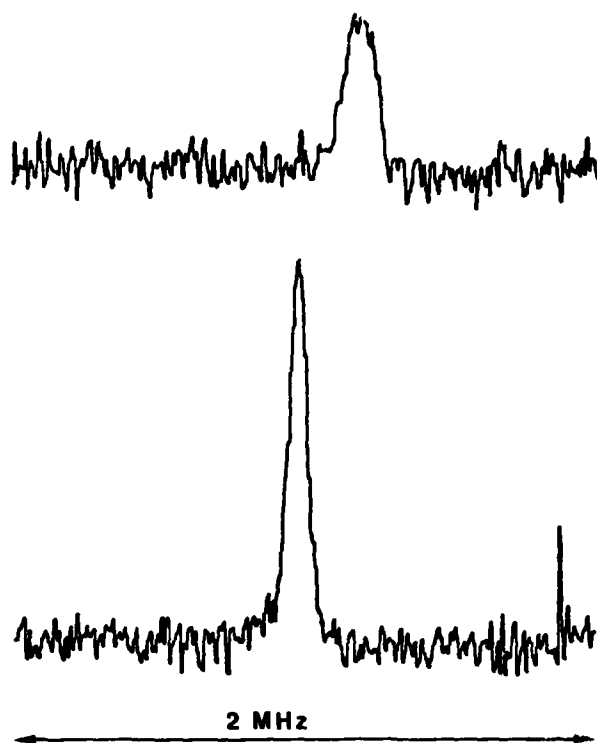


Fig. 3 Analyzer display of the FIR laser frequency shift. The lower trace is the beat note for zero field while the upper trace is for a field of 220 V/cm.

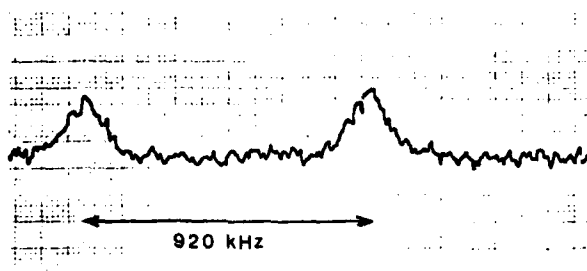


Fig. 2 Spectrum analyzer display of the beat note between the laser and the klystron: an unusual feature which has been occasionally observed is the double peak structure.

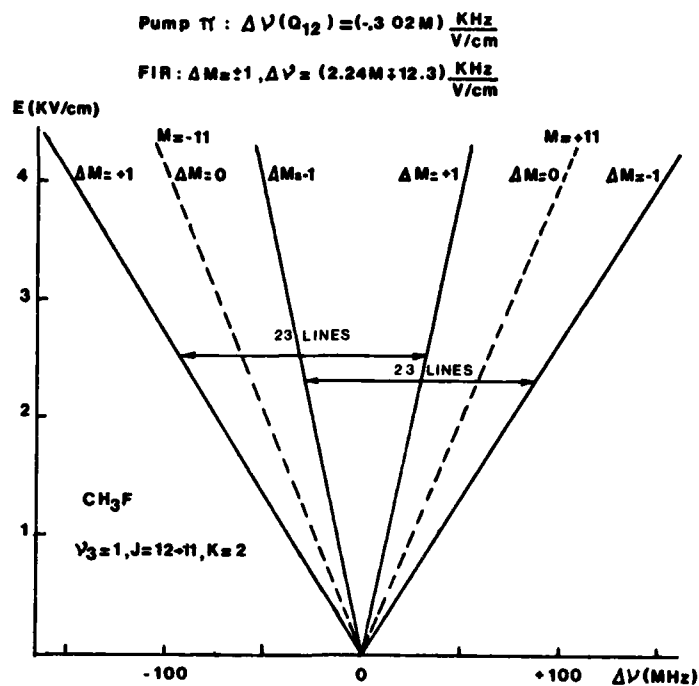


Fig. 4 Frequency shifts of the Stark components of the $496 \mu\text{m}$ line of CH₃F as a function of the applied electric field. Only the lines with $M = \pm 11$ are shown, the others are equally spaced between them.

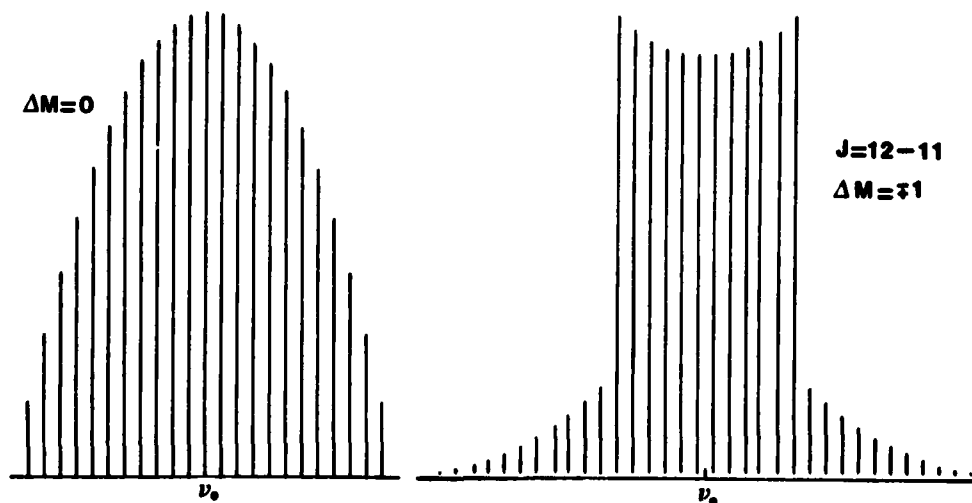


Fig. 5 Intensity pattern of the Stark multiplet for $\Delta M = \pm 1$ and for $\Delta M = 0$ selection rules in the case of a $J = 11-12$ transition.

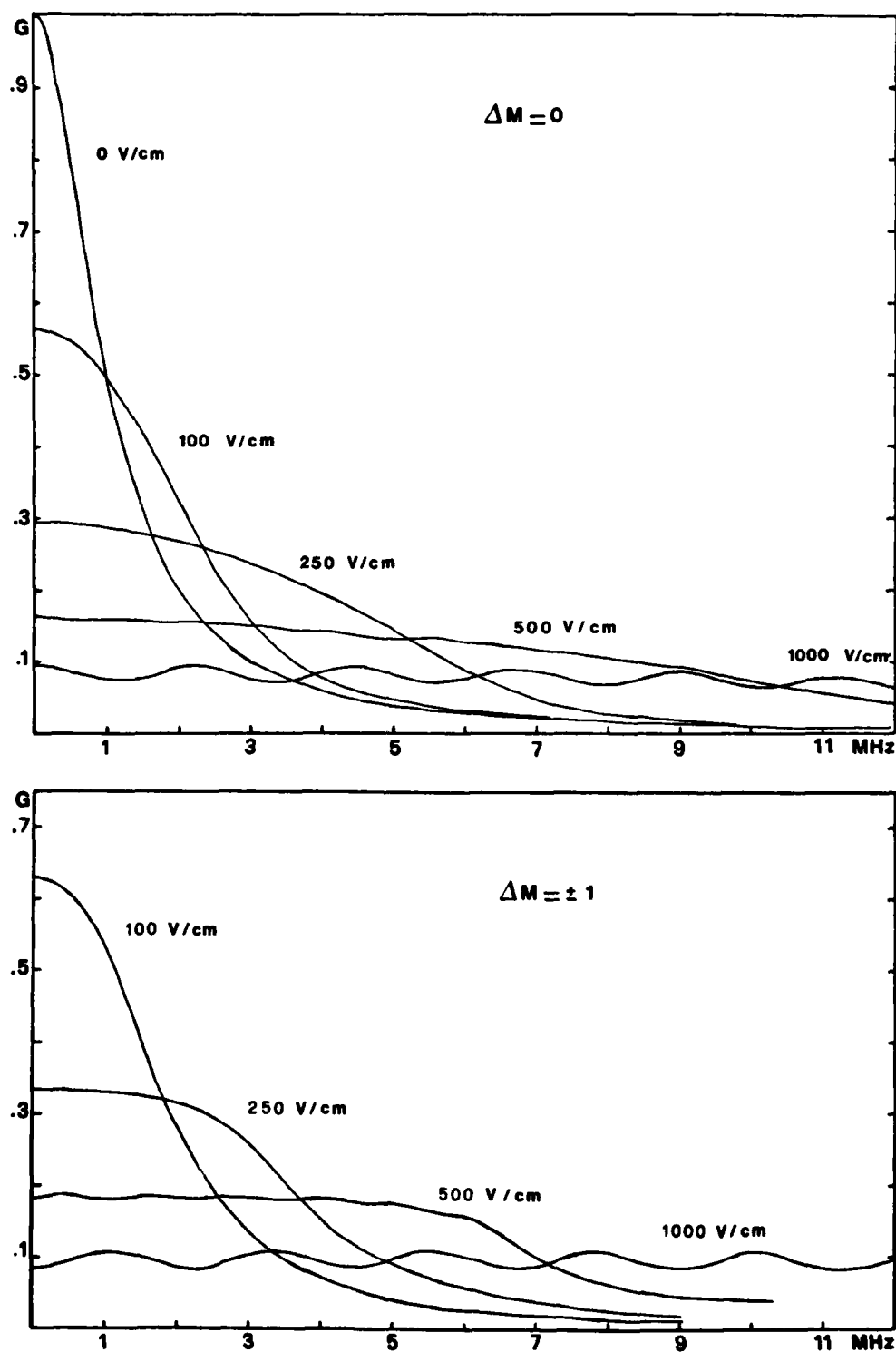


Fig. 6 Relative gain profile of the $496 \mu\text{m}$ line for different values of the static electric field. The gain is normalized to 1 for the maximum without field. All the gain profiles are symmetric around the unperturbed center frequency ν_0 and here only one side is shown. Each single Stark component is supposed to be Lorentzian in shape with a full width of 2 MHz.

THE TRAPPED MERCURY ION FREQUENCY STANDARD

Michael D. McGuire
NASA/Goddard Space Flight Center
Greenbelt, Maryland 20771

Summary

Singly ionized mercury atoms have a structure similar to neutral alkali atoms. They can be maintained as ions for very long times in an rf quadrupole ion trap. Thus their ground state hyperfine structure can be used to make a frequency standard using optical pumping techniques similar to the well known rubidium standard. The mass 199 isotope of mercury has an ionic hfs of 40.5 GHz. In a trap system a linewidth of 10 Hz has been measured. Short term stability of a proposed mercury standard as set by the achieved signal to noise ratio is

$$\Delta f/f = 6 \times 10^{-12} / \sqrt{\tau}.$$

There is prospect of further improvement. Long term stability is affected by second order doppler effect, and by pressure, light, and Stark shifts. However these appear either sufficiently small or sufficiently controllable that the proposed mercury ion standard would be competitive with existing standards.

Introduction

The hyperfine structure of alkali atoms has been the basis of most successful atomic frequency standards. A whole new set of alkali-like "atoms" can be created by singly ionizing the alkaline earth elements such as strontium, barium, or mercury. In spectroscopic language the $1S_0$ ground states go to $2S_{1/2}$ ground states in the ions. The mass 199 isotope of mercury has a nuclear spin of 1/2 and thus a hyperfine structure (hfs) just the same as the simple hfs of hydrogen except that the splitting is 40.5 GHz instead of 1.4 GHz.

An rf quadrupole ion trap will hold ions for hours in a high vacuum well isolated against recombination and other perturbing influences. The trap itself is an arrangement of rf and dc electric fields which create a potential well. The hfs of $3He^+$ was investigated by Schuessler, Fortson, and Dehmelt¹ using a trap technique. Dehmelt² has done a detailed discussion of the theory and operation of ion traps.

Principle of Operation

A frequency standard using the hfs of $199Hg^+$ stored in a trap can be realized using an optical pumping arrangement similar to a rubidium standard. The pump light is from mass 202 isotope mercury ions. This is an even-even isotope and has therefore no nuclear spin and no hfs. A fortuitous consequence of isotope shift is that the $6S_{1/2} \rightarrow 6P_{1/2}$ transition at 194.2 nm in $202Hg^+$ has a good overlap with the same transition in $199Hg^+$ from the $F = 1$ hyperfine substate, but not from the $F = 0$ substate. Figure 1 shows the level structures. The $199Hg^+$ ions excited by the pump light to the $6P_{1/2}$ state fall back to both hyperfine substates of the $6S_{1/2}$ ground state, emitting 194.2 nm fluorescence radiation. In the absence of the 40.5 GHz excitation the $F = 1$ substate is depopulated and the fluorescence disappears. In the literature this is often called intensity pumping. Thus if the hyperfine excitation is present and swept in frequency, the hfs resonance can be seen in the intensity variation of the fluorescence light. This light signal can be used in the usual way to control a frequency

standard oscillator. For comparison in the rubidium optical pumping standard the same sort of set of atomic transitions is used, but absorption of pump light is monitored rather than fluorescence because a much higher particle density is possible in a rubidium cell.

The first work was done with the mercury ion system by Major and Werth³. They measured a linewidth of about 10 Hz. This amounts to a line Q of 4×10^9 . This is matched only by the proposed passive H-maser⁴. Further work was done with the mercury ion system by McGuire, Petsch, and Werth⁵. They achieved a large improvement in signal to noise ratio that now makes it possible to consider building a frequency standard using the mercury ion.

Apparatus

The apparatus of McGuire et al. is shown schematically in figure 2. The trap electrodes are conjugate hyperboloids of revolution. A characteristic dimension is z_0 , half the distance between the two sheet parts, about 15 mm. The trap holds $199Hg^+$ ions for times ranging to hours. A $202Hg^+$ discharge lamp excited by about 100 watts of 100 MHz rf provides the pump light. The method of fabrication and the operating conditions strongly affect its output at 194.2 nm. The light from the lamp passes through the trap and is monitored. The monitor signal can be used to reduce lamp noise. The fluorescence signal is taken at right angles to the incoming light to avoid being swamped by the incoming light since it is all at the same wavelength. Herein lies the greatest part of the signal to noise problem. Scattered and reflected light from the incoming beam forms a large background to the fluorescence signal. To date the best compromise between a large amount of pump light reaching the trapped ions and sufficient collimation to cut down the reflection and scattering has resulted in a signal about 2.5% of background.

The 40.5 GHz excitation is obtained from a klystron oscillator phase locked to a harmonic of a low frequency synthesizer of good spectral purity and stability. The electron gun creates ions in the trap from the low pressure background of mercury atoms allowed to exist in the trap vacuum chamber. A solar blind (sensitive to only uv light) photomultiplier is used to observe the fluorescence signal. An interference filter peaked at 194 nm removes light from the electron gun and also discriminates against a strong atomic line of mercury at 254 nm that is present in the background. The photomultiplier signal is processed in photon counting mode to obtain best possible noise performance at the low signal level. A frequency standard would exist if the fluorescence signal were fed back to control the frequency of the klystron—a frequency locked loop.

Short Term Stability

The background count rate was about 250,000/sec and the signal was about 2.5% of this. Assuming square-wave frequency modulation is used between the two steepest points on opposite sides of the Lorentzian resonance curve to obtain the discriminator signal, we obtain a shot noise limited short term frequency stability of

$$\delta f/f = 6.4 \times 10^{-12}/\sqrt{\tau}$$

as illustrated in figure 3. It is comparable to existing cesium and rubidium standards.

Further improvement in signal to noise ratio is in prospect. Considerable reduction of light background should be possible with a right circular cylinder (tuna fish can) trap instead of a hyperboloidal trap. The idea is illustrated in figure 4. An analysis by Benilan and Audoin⁶ indicates minimal change in trapping behavior. Werth and Ifflaender⁷ achieved complete suppression of the background in an experiment with trapped barium ions by using a laser. Unfortunately a laser at 194.2 nm is not immediately in prospect for the mercury ion standard. However the limits have probably not been reached for improvement in lamp intensity. A factor of two or three improvement may lurk in the overall quantum efficiency of the photomultiplier--interference filter combination.

Systematic Effects and Long Term Stability

The fact that the ions are harmonically bound in an active potential well has various consequences in terms of frequency shifts that are produced. The first order doppler shift is nil. The Dicke⁸ theorem predicts that the first order spectrum consists of the unshifted central line plus sidebands spaced at integer multiples of the harmonic motional frequency (typically 50 KHz) away from the central line. The buffer gas has a similar effect in a rubidium cell.

The second order doppler shift is of far greater consequence. It has been predicted² and found experimentally⁷ that the average energy of ions in an rf trap is about 1/10 of the well depth. The ratio of fractional shift to kinetic energy for mercury ions is $5 \times 10^{-12}/\text{eV}$. Experimental results are shown in figure 5. The error bars reflect limitations in the spectral purity of the 40.5 GHz source. A deep potential well is desirable to trap a greater density of ions for a higher signal to noise ratio while a shallow well would minimize the shift. However a stabilized shift is acceptable for frequency standard purposes. If everything scales, 0.1% stabilization of the trap voltages would bring fractional fluctuations in the shifted frequency to the 10^{-15} region. Dehmelt² has suggested using a buffer gas of light atoms such as helium to reduce the second order doppler shift by collisional cooling. This would be at cost of some pressure shift, not present in the usual 10^{-10} operating vacuum for the trap system. The pressure shift for mercury ions in helium is not known, but Vetter, Stuke, and Weber⁹ measured the shift for structurally very similar barium ions in helium at about 10^2 Hz/torr. A 10^{-6} torr buffer gas pressure might produce useful cooling of the mercury ions. Thus the fractional pressure shift would be in the 10^{-13} region. A pressure--temperature stabilization scheme could reduce fluctuations in the shift by two more orders of magnitude.

The electric fields of the trap itself might be expected to produce some Stark effect shift. These fields are of the order of 500 V/cm. To lowest order the Stark shift of hydrogen-like hfs varies as E^2 . No measurements are available for a trap standard. Ruan and Zacharias measured the shift for cesium and found a $\delta f/f = 2.5 \times 10^{-16} E^2$, (E , V/cm) which in a trap would amount to a $\delta f/f = 6 \times 10^{-11}$. However ions have smaller polarizability than isoelectronic atoms. Also fields at the center of the trap are smaller and vary quadratically outwards. The ions remain mostly in the low field center of the trap. Assuming the ions remain in a spherical volume of radius half the maximum of the

trap, the field is down by a factor of four and the shift by a factor of sixteen. If the ionic polarizability is 1/10 that of the isoelectronic atom, then the shift would be $\delta f/f = 4 \times 10^{-14}$. The stabilization of trap voltages to control the second order doppler shift would also stabilize this shift, probably into the 10^{-17} region.

Hyperfine structure is susceptible to magnetic shifts. We use the ($F = 1, m = 0$) to ($F = 0, m = 0$) transition which has quadratic field dependence. From the Breit-Rabi formula

$$\delta f = (g_J - g_I)\mu_B H^2 / 2h\nu_{\text{hfs}}$$

where g_J and g_I are the electronic and nuclear g factors, μ_B is the Bohr magneton and ν_{hfs} is in this case 40.5 GHz. For $^{199}\text{Hg}^+$ we have

$$\delta f = 97.2 \text{ Hz}^2, H \text{ in gauss}$$

The high hyperfine frequency makes this the smallest magnetic shift of all the atomic frequency standards so that the shielding and magnetic construction material problem is slightly alleviated. Figure 6 shows some measurement results.

Because there is not perfect overlap between the pump light spectrum and the absorption spectrum of the trapped ions (see figure 1) there will be a linear shift in the hyperfine frequency with light intensity. Data on the relative spectral profiles and offsets of sufficient accuracy are not at hand. A similar problem exists with the rubidium standard. Possible solutions are isotope mixtures in the pump lamps or alternate application of microwave and light excitation to the ions¹¹.

Spin exchange shifts will be minimal at the low background pressures used. This a basic advantage of electrodynamic confinement of ions over physical confinement of atoms with a buffer gas or a coated cell.

Conclusions

In summary the achievable short term stability and prospects for further improvement make it attractive to continue further development of the mercury ion standard. The estimates of long term shifts and sources of error are not discouraging, but are generally speculative and based mostly on other systems. As is true of other frequency standards, some short term stability will have to be traded off for improvement in long term stability. A large part of the work to be done lies in obtaining a good bargain in this respect.

References

1. H. Schuessler, N. Fortson, and H. Dehmelt, Phys. Rev. **187**, 5 (1969)
2. H. Dehmelt, Advances in Atomic and Molecular Physics, **3**, 53 (1969) and **5**, 109 (1969)
3. F. Major and G. Werth, Phys. Rev. Lett. **30**, 1155 (1973)
4. F. Walls and H. Hellwig, Proceedings of the 30th Annual Symposium on Frequency Control, 473 (1976)
5. M. McGuire, R. Petsch, and G. Werth, Abstracts, Fifth Int. Conf. on At. Phys., Berkeley 1976, p407
6. M-N. Benilan and C. Audoin, Int. J. Mass Spectr. and Ion Phys. **11**, 421 (1973)

7. R. Ifflaender, and G. Werth, Abstracts, Fifth Int. Conf. on At. Phys., Berkeley 1976, p404
8. R. H. Dicke, Phys. Rev. 89, 472 (1953)
9. J. Vetter, M. Stuke, and E. Weber, Z. Physik A273, 129 (1975)
10. R. Huan and J. Zacharias, Phys. Rev. 107, 107 (1957)
11. M. Arditi and T. Carver, Phys. Rev. 124, 800 (1961)

Layout of Mercury Ion Frequency Standard

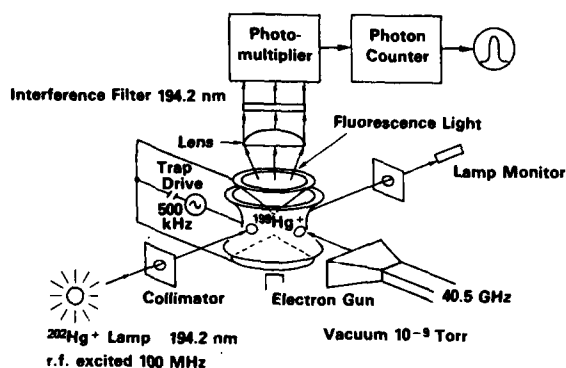


FIGURE 2

Right Circular Cylinder Trap

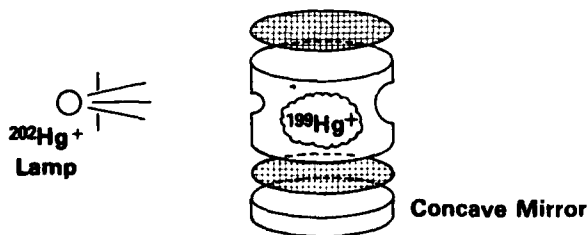


FIGURE 4. Right circular cylinder (Tuna Fish Can) trap.

Optical Pumping Scheme for Mercury Ion Frequency Standard

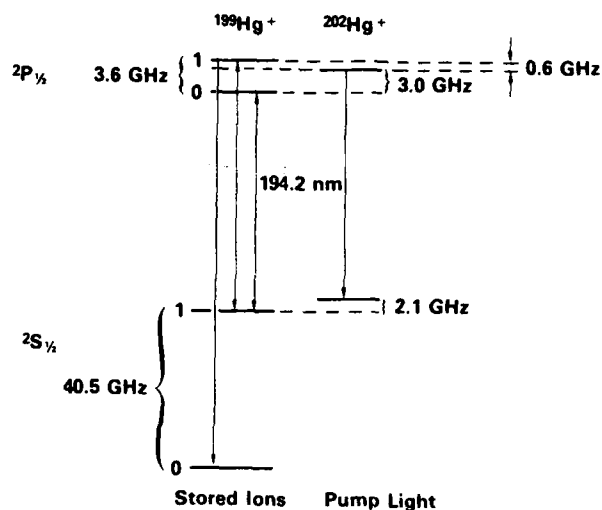


FIGURE 1

CALCULATED STABILITY OF MERCURY STANDARD

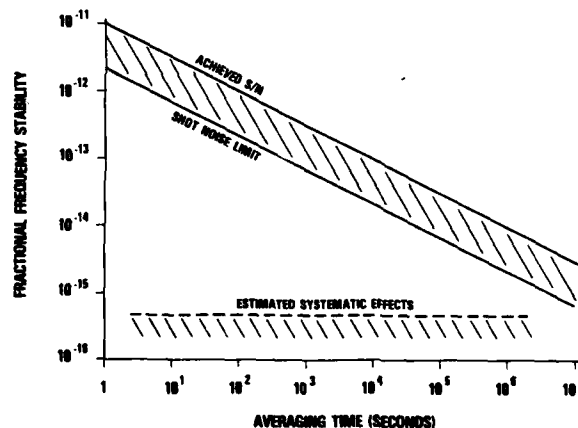


FIGURE 3. Predicted short term stability of mercury Ion frequency standard.

Shift of the Hyperfine Resonance of $^{199}\text{Hg}^+$ as a Function of the Potential Well Depth of an RF Trap

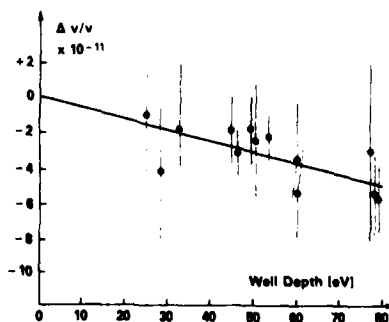


FIGURE 5

Magnetic Field Dependence of the $\Delta F = 1$,
 $\Delta m = 0$ Hyperfine Structure Transition in $^{199}\text{Hg}^+$

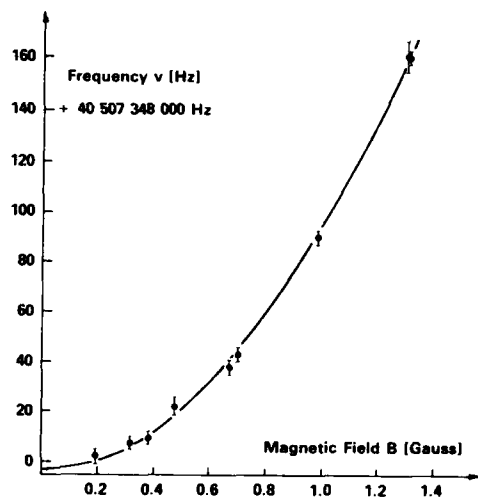


FIGURE 6

$\Delta F = 1, \Delta m_F = 0$ Transition in $^{199}\text{Hg}^+$

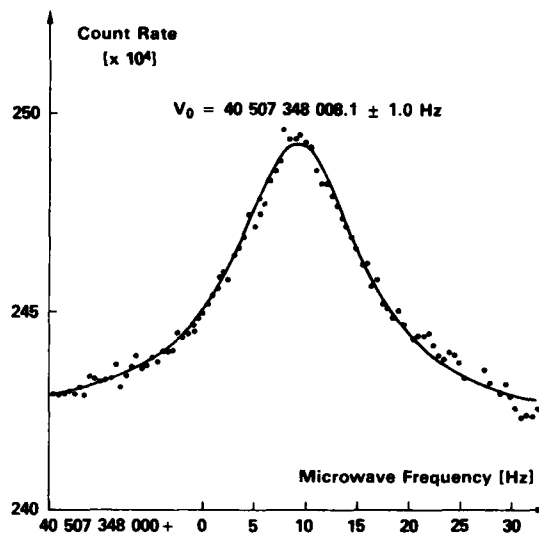


FIGURE 7. Mercury Ion hyperfine resonance, 10 sec/
 channel averaging time.

CLOCKS BASED UPON HIGH MECHANICAL Q SINGLE CRYSTALS*

D. F. McGuigan and D. H. Douglass
Department of Physics & Astronomy
University of Rochester
Rochester, New York 14627

Summary

Large, nearly perfect single crystals of silicon and sapphire are commercially available with masses of many kilograms. These crystals have been shown to have very large mechanical quality factors at low temperatures when oscillating in low mechanical modes.¹ In this paper we consider the possibility of making these mechanical oscillators into stable clocks and frequency standards. The fundamental limits to frequency stability arising from the Brownian motion of the oscillator, the electrical noise, and the quantum limit are all considered. All of these limits can, with proper choice of parameters, be below the performance levels of presently available frequency sources. The factors which limit the frequency stability are likely to be due to time varying processes within the crystal.

Key words: Mechanical oscillator, frequency stability, silicon single crystal, High Q, Quality factor, Brownian motion limit, and quantum limit.

Introduction

In connection with our efforts to construct very sensitive detectors for gravitational radiation we have been studying the properties of large nearly perfect single crystals of silicon and sapphire. These crystals are cylinders of mass several kilograms instrumented to oscillate in their first longitudinal mode of vibration, typically 20 kilohertz, at temperatures the order of a few degrees Kelvin. One of the important parameters for good gravitational detectors is a high mechanical quality factor, Q , which is related to the amplitude relaxation time, τ^* , by

$$Q = \frac{\omega \tau^*}{2} \quad (1)$$

where ω is the (angular) frequency of vibration. We report here Q values in excess of 10^9 .

After a little reflection one realizes that these gravitational wave detectors can easily be made into clocks. Furthermore, with Q values this high one might expect to make a very good clock. How good these clocks can be is the subject of this paper.

Experimental Methods and Q Measurements

Figure 1 illustrates the experimental apparatus. A crystal cylinder which vibrates in its first longitudinal mode of vibration is suspended against gravity in a vacuum chamber by a tungsten wire (typically 0.025 cm diam). The wire is located at the middle as shown, which is a node for this mode of vibration. The wire is attached to a "cold plate" which can be at any temperature above 1.4 K. The temperature of the crystal can be controlled by changing the temperature of the "cold plate." Liquid helium, which is the source of the cooling, surrounds the vacuum chamber in a standard cryostat. Figure 2 shows schematically the electrical circuit which we are presently using for exciting the crystal into resonance and the detection of the subsequent motion. This scheme is essentially that of a capacitance microphone. In this circuit the displacement ΔX of the face of the crystal causes a change in the capacitance C , which produces a change in the output signal ΔV at the resonator frequency ω

$$\Delta V = V_A - V_B \quad (2)$$

The voltages V_A and V_B are put into the differential

input of a lock-in amplifier and $V_A - V_B$ is demodulated with a local oscillator [a Rockland 5100 frequency synthesizer locked to a high stability quartz oscillator] at a frequency close to ω . If $RC_1 \gg \frac{2\pi}{\omega}$ (where $2\pi/\omega$ is the period of oscillation of the crystal), the relationship between ΔX of the crystal and the signal ΔV at ω is

$$\Delta V \approx \frac{V}{C} \left(\frac{\partial C}{\partial X} \right) \Delta X \quad (3)$$

The value of $\frac{1}{C} \frac{\partial C}{\partial X}$ was approximately 10 to 100 cm⁻¹. The precise value depends on the electrical conductivity of the faces of the crystal. For the voltages used in this experiment, $V_0 \sim 10$ to 200 volts, $V \sim 10^{-7}$ to 10^{-3} volts, the range of displacements measured was 5 x 10^{-11} to 10^{-6} cm.

Figure 3 shows the amplitude of vibration versus time of a silicon cylinder² (mass 4.9 kg, length 22.9 cm, diameter 10.6 cm, $f(4.2 \text{ K}) = 19,553.46542 \text{ Hz}$) in free decay at a temperature of 3.5 K. The measured relaxation time is 3.1×10^8 sec., which corresponds to a Q of 2×10^9 .

If one had a crystal free from impurities and imperfections one might expect significantly higher Q 's. We note that the Q values reported here are comparable in magnitude to the best values reported in other pure materials: 4.2×10^9 in quartz³ and 5×10^9 in sapphire.⁴ The question of how large Q can be in a pure crystal is an interesting one. The ultimate loss mechanism is due to phonon-phonon scattering and an estimate of the magnitude leads to very large Q values at low temperatures. Before reaching this limit one will probably encounter the case of a pure crystal with rough surfaces. Savedoff⁵ has given an argument to suggest that for this case $Q \sim (\pi L/a)^2$, where L is the size of the crystal and a is a measure of the surface roughness. Choosing $L \sim 30$ cm, $a \sim 10^{-7}$ cm, one estimates that $Q \sim 10^{18}$ is feasible, which is seen to be 8 or 9 orders of magnitude higher than what we have presently achieved.

Frequency Stability of a Mechanical Oscillator

We consider a mechanical oscillator of mass M vibrating with amplitude x at angular frequency ω . The uncertainty in the determination of the frequency f is related to the uncertainty in determining the time interval τ of an integral number of cycles:

$$\frac{\Delta f}{f} \approx \frac{\Delta \tau}{\tau} \quad (4)$$

The uncertainty in τ is caused by a fluctuation in the amplitude

$$\Delta \tau \approx \frac{\partial \tau}{\partial x} \Delta x = \frac{1}{\omega x} \Delta x \quad (5)$$

so that we have

$$\frac{\Delta f}{f} \approx \frac{1}{\omega \tau x} \Delta x \quad (6)$$

The energy of the oscillator is

$$E \approx M \omega^2 x^2 \quad (7)$$

The uncertainty Δx in equation 6 can be caused by statistical fluctuations and is

$$\Delta x = (\overline{\Delta x_1^2} + \overline{\Delta x_2^2} + \dots)^{1/2} \quad (8)$$

where the bar signifies a statistical average and where allowance for more than one source has been explicitly indicated. Equation (6) can be expressed as

$$\frac{\Delta f}{f} = \frac{1}{\omega \tau} \left(\frac{\Delta E}{E} \right)^{1/2}, \quad (9)$$

where equation (7) for E has been used and where

$$\begin{aligned} \Delta E &= M\omega^2 (\overline{\Delta x_1^2} + \overline{\Delta x_2^2} + \dots), \\ &= \Delta E_1 + \Delta E_2 + \dots \end{aligned} \quad (10)$$

has been defined. Equation (9) is the basic relationship. The fluctuations in the frequency can be determined by identifying the various sources of uncertainty in energy as expressed by equation (10).

Brownian Motion of the Oscillator

The fluctuation in the amplitude Δx of an oscillator at temperature T due to the Brownian motion is easily worked out⁶ and is

$$\overline{\Delta x^2}_{BM} = \frac{n k T}{M \omega^2} \quad (11)$$

where k is Boltzman's constant and

$$\eta = \begin{cases} \frac{2\tau}{\tau^*} & , \quad \tau \ll \tau^* \\ 1 & , \quad \tau \geq \tau^* \end{cases} \quad (12ab)$$

and where τ^* is the amplitude relaxation time given in equation (1). Because of our high Q's and large τ^* 's we will consider only case (12a). Putting equation (11) into (9) we obtain

$$\left. \frac{\Delta f}{f} \right|_{BM} \approx \frac{1}{\omega \tau} \left(\frac{2\tau}{\tau^*} \frac{kT}{E} \right)^{1/2} \quad (13)$$

Using equation (1) this can be expressed

$$\left. \frac{\Delta f}{f} \right|_{BM} \approx \left(\frac{kT}{\tau Q \omega E} \right)^{1/2} \quad (14)$$

If one remembers a relation between the energy loss \dot{E} of the oscillator and the quality factor Q

$$Q = \frac{\omega E}{\dot{E}}, \quad (15)$$

then (14) can be put into a more familiar form

$$\left. \frac{\Delta f}{f} \right|_{BM} \approx \left(\frac{kT}{E Q^2 \tau} \right)^{1/2} \quad (16)$$

If the power P delivered to the measuring circuit is one-half of the power lost as it frequently is in an optimum impedance match, equation (16) becomes

$$\left. \frac{\Delta f}{f} \right|_{BM} \approx \left(\frac{kT}{2P Q^2 \tau} \right)^{1/2}, \quad (17)$$

which is identical to what one finds in the literature.⁷ We shall work with the fluctuation in f as expressed by equation (14). Putting the expression for the energy, equation (7), into (14) we obtain

$$\left. \frac{\Delta f}{f} \right|_{BM} \approx \left(\frac{kT}{\tau M \omega^2 x^2 Q} \right)^{1/2} \quad (18)$$

We will now estimate the order of magnitude of Δf . Let us assume $M = 1 \times 10^{-6}$ gms, $T = 1K$, $\omega = 2\pi \times 2 \cdot 10^4$ sec⁻¹, $x = 10^{-4}$ cm, $Q = 10^3$, which corresponds to values of parameters that can be easily achieved.

$$\left. \frac{\Delta f}{f} \right|_{BM} = 8.4 \times 10^{-19} \left(\frac{10^{-4} \text{ cm}}{x} \right) \left(\frac{2 \times 10^4 \text{ Hz}}{f} \right)^2 \left\{ \left(\frac{1 \text{ sec}}{\tau} \right) \left(\frac{T}{1K} \right) \left(\frac{10^{-6} \text{ gms}}{M} \right) \left(\frac{10^3}{Q} \right) \right\}^{1/2} \quad (19)$$

The Q dependence of this expression is shown in Figure 4 for integration times of 1 sec and 100 sec. One sees that the frequency fluctuations from this source are easily smaller than that of available sources (such as the Hydrogen maser or the superconducting cavity-stabilized oscillator). In principle one can make this source of uncertainty as small as we desire by making T smaller and/or making oscillators with larger Q's.

Electrical Noise. Let us suppose that we can make the frequency fluctuation due to the Brownian motion as small as we like. At some point we will encounter the noise in the transducer and amplifiers in the electrical circuit. Thus there will be an energy uncertainty ΔE_e which when determined can be put into equation (9) to find the frequency uncertainty. Braginsky⁸ and Giffard⁹ have shown that ΔE_e of a transducer-amplifier system is of the form

$$\Delta E_e = A \frac{1}{\tau} + B \tau, \quad (20)$$

where A and B are parameters describing the particular transducer-amplifier scheme. The first term is frequently associated with the noise from the amplifier and the second term is frequently referred to as the Braginsky back-reaction effect. It is sometimes convenient to define a noise temperature T_N

$$\Delta E_e = k T_N. \quad (21)$$

We note that at short integration times the first term dominates, giving a $\Delta f/f \sim \tau^{-1/2}$ dependence, and at long integration times the second term dominates, yielding $\Delta f/f \sim \tau^{1/2}$. Let us consider the frequency fluctuation using expression (21)

$$\left. \frac{\Delta f}{f} \right|_e \approx \frac{1}{\omega \tau} \left(\frac{k T_N}{E} \right)^{1/2} \quad (22)$$

Using the expression for E and the values of the parameters used previously we obtain

$$\left. \frac{\Delta f}{f} \right|_e = 7.4 \times 10^{-17} \left(\frac{10^{-4} \text{ cm}}{x} \right) \left(\frac{2 \times 10^4 \text{ Hz}}{f} \right)^2 \left\{ \left(\frac{1 \text{ sec}}{\tau} \right) \left(\frac{T_N}{1K} \right) \left(\frac{10^{-6} \text{ g}}{M} \right) \right\}^{1/2} \quad (23)$$

To estimate the order of magnitude of $\Delta f/f$ we consider several cases.

FET Amplifier. One of the best available room-temperature amplifiers is made from Field Effect Transistors (FET). Although T_N varies with τ , as equation (20) shows, and also with circuit impedance, we will estimate the frequency fluctuation assuming $T_N \sim 0.25$ K, which is close to what can be achieved. For $\tau \sim 1$ sec,

$$\left. \frac{\Delta f}{f} \right|_{FET} \sim 3.7 \times 10^{-17}, \quad (24)$$

which is competitive with the best available frequency standards. One could wish to do better. It is possible that one could drive the oscillator at a larger amplitude, say 10^{-2} cm. At some point, though, one would expect to encounter nonlinear effects which would cause frequency shifts.

Low Temperature Transducers. Since one has to

cool these crystals to achieve the high Q's, one might wish at the same time to use the new transducers being developed for use with the gravitational wave detectors. Some of the most sensitive ones being proposed are based upon the SQUID (Superconducting QUantum Interference Device). The present state of the art suggests that one could build a transducer amplifier with $T_N \sim 10^{-5}K$ even though there are severe problems such as impedance matching, which must be overcome. Putting this value into equation (23) one obtains

$$\left. \frac{\Delta f}{f} \right|_{\text{SQUID}} \sim 2.3 \times 10^{-19} \left(\frac{10^{-4} \text{ cm}}{x} \right) \left(\frac{2 \times 10^4 \text{ Hz}}{f} \right)^2 \cdot \left(\frac{1 \text{ sec}}{\tau} \right) \left(\frac{10^4 \text{ gm}}{M} \right)^{\frac{1}{2}} \quad (25)$$

Thus we see that if one could reach the room temperature "limit" of 10^{-17} given by equation (24), then two more orders of magnitude could be achieved by going to the more exotic transducers being developed for the gravitational wave detectors.

The Quantum Limit. One can reasonably expect that the low temperature transducers being developed for the gravitational wave detectors will get better as more work is spent on them. Braginsky and Giffard have also shown that there is a fundamental limit ΔE_q for linear devices on the value ΔE_e which comes from the uncertainty principle in quantum mechanics

$$\Delta E_q \geq \Delta E_e \approx 2\hbar\omega \quad (26)$$

where \hbar is Planck's constant. Let us put this in equation (9)

$$\left. \frac{\Delta f}{f} \right|_q \approx \left(\frac{2\hbar}{M\tau^2\omega^3 x^2} \right)^{\frac{1}{2}} \quad (27)$$

Using the same example as before

$$\left. \frac{\Delta f}{f} \right|_q \sim 1.0 \times 10^{-19} \left(\frac{10^{-4} \text{ cm}}{x} \right) \left(\frac{2 \times 10^4 \text{ Hz}}{f} \right)^2 \cdot \left(\frac{1 \text{ sec}}{\tau} \right) \left(\frac{10^4 \text{ gm}}{M} \right)^{\frac{1}{2}} \quad (28)$$

The example of the SQUID, equation (25), is close to this limit. This is the ultimate limit for a linear detector. However, Braginsky⁶ has proposed a "quantum non-demolition" scheme which may not be subject to this limitation.

Experimental Plan

Our plan is to study these large nearly perfect single crystals as stable frequency sources. One of the first experiments will be to compare two of these oscillators using room-temperature electronics. The circuit of Figure 2 is analyzed in the appendix. There we show that the Johnson noise from the resistor dominates with the values presently being used. It should be relatively easy, however, to reduce the Johnson noise below the limit of the best FET amplifiers. Next we want to compare one of these oscillators with one of the primary standards, a Cs beam clock say. This would allow us to study the time dependence of effects within the crystal that would cause the crystal to change its length or to change its elastic constants and hence its frequency. It is these effects that we expect may limit the frequency stability of the crystal. Other effects would also be studied, such as temperature changes (we expect this to be an easy problem since the change in frequency with temperature is small; $f-f(4.2 K) \sim 1.8 \times 10^{-10} T^3$ for the crystal described in this paper.)

If we do reach the limit of room-temperature elec-

tronics we will then consider using the more exotic low-temperature transducers such as the SQUID.

Appendix A Analysis of Circuit in Figure 2

As an example of the frequency fluctuations which occur in a particular scheme we consider the one in Figure 2. The relationship between the displacement Δx and the change in the voltage ΔV is given by equation (3)

$$\Delta V \approx V_o \left(\frac{1}{c} \frac{\partial c}{\partial x} \right) \Delta x \quad (A1)$$

We can calculate the frequency fluctuation for this case directly from equation (6)

$$\frac{\Delta f}{f} \approx \frac{\Delta V}{\omega \tau x V_o \left(\frac{1}{c} \frac{\partial c}{\partial x} \right)} \quad (A2)$$

We now assume that we are using an amplifier with a noise voltage V_n given by

$$(V_n^2)^{\frac{1}{2}} = (4kT_a R_a \Delta f)^{\frac{1}{2}} \quad (A3)$$

where there is a characteristic temperature T_a and characteristic resistance R_a describing the amplifier. If we assume that we are limited by this noise voltage, then the frequency fluctuation can be expressed as

$$\frac{\Delta f}{f} \approx \frac{1}{\omega \tau x \left(\frac{1}{c} \frac{\partial c}{\partial x} \right) V_o} (4kT_a R_a \Delta f)^{\frac{1}{2}} \quad (A4)$$

Next we assume that we have a very good FET amplifier (with $4kT_a R_a \sim 10^{-19}$ volts $\text{Hz}^{-\frac{1}{2}}$). Using parameters that we have discussed above we obtain

$$\frac{\Delta f}{f} \sim 2.5 \times 10^{-15} \left(\frac{2 \times 10^4 \text{ Hz}}{f} \right) \left(\frac{10^{-4} \text{ cm}}{x} \right) \left(\frac{100 \text{ cm}^{-1}}{c} \frac{\partial c}{\partial x} \right) \cdot \left(\frac{100 \text{ V}}{V_o} \right) \left(\frac{4kT_a R_a}{10^{-19} \text{ V}^2 \text{ Hz}^{-1}} \right)^{\frac{1}{2}} \left(\frac{1 \text{ sec}}{\tau} \right)^{\frac{1}{2}} \quad (A5)$$

where we have taken account of the fact that $\Delta f \approx \frac{1}{\tau}$. This is of the same order of magnitude as was achieved above (see equation 23) but here one sees how the various parameters come in. We mention again that the $(4kT_a R_a)^{\frac{1}{2}} \sim 3 \times 10^{-10}$ volts $\text{Hz}^{-\frac{1}{2}}$ value can only be achieved at a certain frequency and impedance and if this can not be achieved then one will of course do worse than that given by equation A5.

In order to reach the limit given by (A5) one has to make sure that the Johnson noise from the external resistor is low enough. The noise voltage V_R appearing across C in Figure 2 is

$$V_R \approx \frac{1}{\omega RC} (4kTR\Delta f)^{\frac{1}{2}} \quad (A6)$$

where R and C are the resistance and capacitance indicated in the measuring circuit. This can be expressed as

$$\frac{V_R}{\sqrt{\Delta f}} = 1.9 \times 10^{-9} \left\{ \left(\frac{T}{K} \right) \left(\frac{10^7 \Omega}{R} \right) \left(\frac{10^{-11} \text{ f}}{C} \right)^2 \left(\frac{2 \times 10^4 \text{ Hz}}{f} \right) \right\}^{\frac{1}{2}} \text{ volts } \text{Hz}^{-\frac{1}{2}} \quad (A7)$$

At room temperature and with the parameters presently being used this source of noise would dominate that from a good FET preamplifier (3×10^{-10} volts $\text{Hz}^{-\frac{1}{2}}$). However, one can reduce V_R by lowering T and increasing C and R.

References

*Supported in part by the National Science Foundation

1. D. H. Douglass, "Gravitational Wave Experiments" to be published in the proceedings of the Academia Nazionale dei Lincei International Symposium on Experimental Gravitation, Pavia, Italy. September 17-20, 1976. Also McGuigan, Lam, Gram, Hoffman and Douglass. To be published.
2. This crystal was grown by the zero-dislocation Czochralski process with the [111] direction parallel to the cylinder axis and was obtained from Monsanto Company. The crystal is a p-type semiconductor doped with approximately 4×10^{15} Boron atoms/cc ($\rho \sim 3-5$ ohm-cm). Other impurity atoms are present—for example carbon at the 8×10^{16} atoms/cc level.
3. A. G. Smagin, Cryogenics 15, 483 (August, 1975),
4. V. B. Braginsky, Proceedings of the Academia Nazionale dei Lincei International Symposium on Experimental Gravitation. Pavia, Italy. September 17-20, 1976. See also Sov. Phys. Crystallogr. 19, 549 (1974).

5. M. Savedoff, private communication. A preliminary discussion of the model is given in reference 1 above.
6. V. B. Braginsky and A. B. Manukin, Measurement of Weak Forces in Physics Experiments, ed. D. H. Douglass. University of Chicago Press, 1977.
7. L. S. Cutler and C. L. Searle, Proceedings of the IEEE 54, p. 136. (1966).
8. V. B. Braginsky, Proceedings of the International School of Cosmology and Gravitation. Erice, Italy. March 13-25, 1975.
9. R. P. Giffard, Physical Review-D 14, p. 2478, (1976).

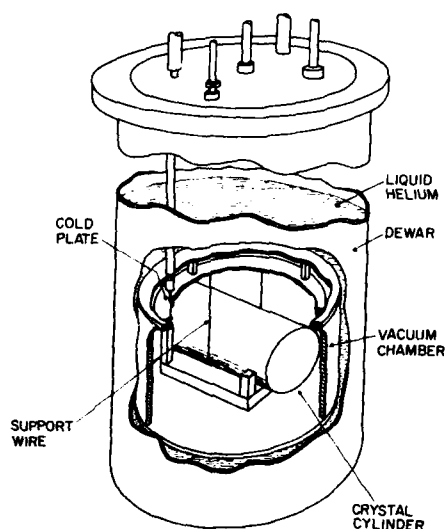


Figure 1: Schematic of crystal cylinder and liquid helium cryostat.

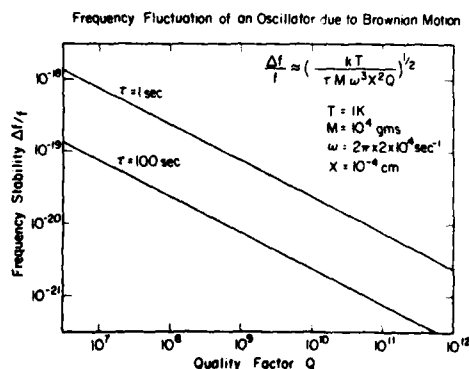
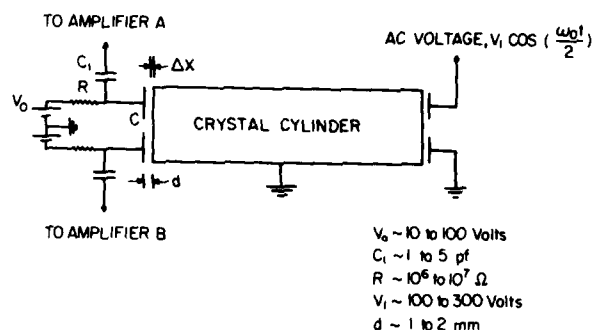


Figure 4: Frequency fluctuation of an oscillator vs. Quality factor due to Brownian Motion. Curves for integration times of 1 sec. and 100 sec. are shown.



SCHEMATIC OF Q MEASURING CIRCUIT

Figure 2: Schematic of electrical circuit. A voltage at one half the cylinder frequency excites the cylinder into resonance. The amplitude of vibration ΔX is determined by measuring the output $V_A - V_B$ of the balanced capacitance microphone circuit. Displacements of $X \sim 5 \times 10^{-11}$ cm can easily be measured.

VIBRATION OF SILICON CYLINDER

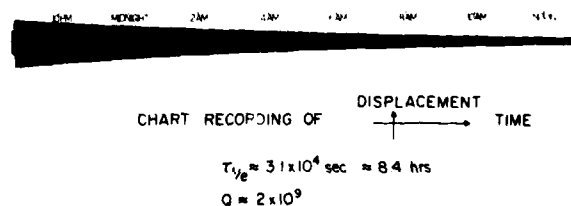


Figure 3: Amplitude of vibration in first longitudinal mode of vibration vs. time. The value of $Q \sim 2 \times 10^9$ corresponding to a relaxation time $(1/e)$ of 3.1×10^4 sec = 8.6 hrs. was measured. The oscillation in amplitude is caused by a slight frequency difference (which is deliberate) between the local oscillator and the cylinder. The frequency at $T = 4.2$ K was 19553.46542 Hz.

INDEX OF AUTHORS

AUTHOR	PAGE	AUTHOR	PAGE	AUTHOR	PAGE
C. A. Adams	3, 246	J. Hermann	55	D. B. Percival	319
D. W. Allan	495, 555	J. Ho	421	H. E. Peters	510
T. L. Anderson	171	D. T. Hodges	574	P. Petit	499, 520, 535
R. Arakelian	400	T. Hoffman	525	J. C. Poncot	17
S. H. Arneson	231	J. D. Holmbeck	390	K. A. Prichard	495
T. Arranz	207	F. W. Hopewood	385	R. C. Rennick	191
C. Audoin	311, 520, 535	D. A. Howe	555, 562	A. S. Risley	583
B. A. Auld	251	C. K. Hruska	159	R. C. Rosenfeld	231
H. S. Babbitt, III	412	U. Hubner	327	M. J. Ross	448
W. H. Bailey	187	N. W. Hudson	347	H. Rossman	385
A. Ballato	8, 102	M. Inguscio	605	S. Rovea	535
E. Bava	578	S. Ishigami	62	J. Rutman	291
R. Benedetti	605	A. R. Janus	281	E. S. Sabisky	510
M. Berte	122	S. Jarvis, Jr.	555	A. Sanchez	590
R. J. Beeson	147	A. Javan	590	T. Sase	62
M. Bloch	421	T. R. Joseph	365	M. Sato	62
E. Blomberg	525	A. Kachelmyer	197	J-Y Savard	535
W. L. Bond	153	R. S. Kagiwada	266	J. S. Schoenwald	240
D. B. Bradley	436	K. Kamada	48	L. Schulzke	96
R. Brousseau	344	P. Kartaschoff	478	P. J. Scott	407
K. E. Brown	465	Y. Kasai	207	W. D. Seal	347
T. I. Browning	258	E. D. Kolb	178	J. H. Sherman, Jr.	108
G. Busca	535	Y. Konno	275	S. Shibayama	275
J. B. Cain	436	J. A. Kusters	3, 153, 246	K. Shirai	207
P. H. Carr	182	A. L. Lance	347	F. Shishido	117
R. N. Castellano	126	K. F. Lau	266	B. K. Sinha	23
A. B. Chase	574	R. A. Laudise	178	W. J. Skudera, Jr.	285
C. M. Chie	302	J. G. Leach	3	W. R. Slabon	455
L. E. Cross	171	J. W. LeBus	131	R. C. Smythe	44
E. Daams	463	P.C.Y. Lee	29	G. L. Snider	78
C. F. Davis, Jr.	590	C. Lemieux	463	M. I. Spellman	436
A. De Marchi	578	P. Lesage	311	E. J. Staples	371
J. De Prins	503	M. W. Levine	525	C. M. Stearns	197
M. Desaintfuscién	520, 535	M. F. Lewis	258	S. R. Stein	335, 495, 601
A. Di Lieto	605	J. C. Lievin	503	F. Strumia	605
D. H. Douglass	616	T. C. Lim	371	R. C. Sundahl	126
M. M. Driscoll	400	W. C. Lindsey	302	H. A. Sunkenberg	448
B. Dubouis	499	K. C. Liu	590	F. Tajika	48
L. Dyal III	281	T. Lukaszek	8	S. W. Tehon	197
G. L. Bybwad	144	T. Luxmore	396	M. Tetu	344
E. P. EerNisse	8	N. Manabe	48	H. F. Tiersten	23, 44
G. Elchinder	590	E. Mattison	525	R. Tilton	102
W. L. Eversole	187	M. McCall	574	M. Tonelli	605
B. T. Fang	489	D. F. McGuigan	616	H. S. Tuan	271
G. Faucheron	499	M. D. McGuire	612	C. T. Van Degrieff	375
R. L. Filler	131	T. R. Meeker	35, 126	H. Van de Stadt	601
R. Fischer	96	F. G. Mendoza	347	J. Vanier	344, 535
C. Freed	592	P. Minguizzi	605	P. K. Verma	463
H. Fukuyo	117	M. B. Mohler	562	R. Vessot	525
K. Furusawa	62	J. P. Muhlbaier	385	J. R. Vig	131
A. Gabry	499	D. E. Newell	71, 396	A. I. Vulcan	542
J. J. Gagnepain	17	R. E. Newnham	171	F. L. Walls	335
W. A. Garber	574	S. Nishiyama	275	B. Walther	484
R. M. Garvey	555	P.P. Nuspl	465	K. L. Wang	266
A. Genis	71	G. Nystrom	525	S. Wanuga	197
J. George	542	R. M. O'Connell	182	R. D. Weglein	225
A. Godone	578	M. Okazaki	48	D. J. Wineland	562
P. F. Godwin, Jr.	78	S. Okuda	207	Kuang-Ming Wu	29
E. P. Graf	484	M. Onoe	48, 62	S. D. Wu	271
T. K. Gregory	551	T. F. O'Shea	231	K. Yakuwa	207
T. J. Grenchik	489	J. Otomo	275	Bing-Hui Yeh	251
F. Guillaume	503	N. Oura	117	K. H. Yen	266
P. L. Ham	187	T. E. Parker	359	H. Yoshida	3
H. L. Hartman	429	G. T. Pearman	191		
H. Hellwig	495, 555	C. Pegeot	17		

620 Blank

SPECIFICATIONS AND STANDARDS GERMANE TO FREQUENCY CONTROL

Institute of Electrical and Electronic Engineers

Order through: Institute of Electrical and
Electronic Engineers
345 E. 47th Street
New York, New York 10017

- 176-1949 Piezoelectric Crystals, Terminology for
(ANSI C83.3-1951) (Reaff 1971) \$ 4.00
- 177-1966 Piezoelectric Vibrators, Definitions
and Methods of Measurement for (ANSI C83.17-
1970) \$ 4.00
- 178-1959 Piezoelectric Crystals, Determination of
the Elastic, Piezoelectric, and Dielectric
Constants of, also, the Electromechanical
Coupling Factor (ANSI C83.23-1960) \$ 3.50
- 179-1961 Piezoelectric Ceramics, Methods of Measure-
ment of (ANSI C83.24-1962) (Reaff 1971) \$ 3.50
- 180-1962 Ferroelectric Crystal Terms, Definitions
of \$ 3.00
- 319-1971 Piezomagnetic Nomenclature \$ 4.00

Electronic Industries Association

Order through: Electronic Industries Assn.
2001 Eye Street, N. W.
Washington, D.C. 20006

- (a) Holders and Sockets
 - RS-192-A, Holder Outlines and Pin Connections
for Quartz Crystal Units. (Standard Dimensions
for older types.) \$ 5.00
 - RS-367, Dimensional and Electrical Character-
istics Defining Receiver Type Sockets.
(Including crystal sockets.) \$15.00
 - RS-417, Crystal Outlines (Standard dimensions
and pin connections for current quartz
crystal units - 1974) \$ 5.80
- (b) Production Tests
 - RS-186-D, Standard Test Methods for
Electronic Component Parts \$ 5.20
- (c) Application Information
 - Components Bulletin #6, Guide for the
Use of Quartz Crystal Units for
Frequency Control \$ 3.60

International Electrotechnical Commission

Order through: American National Standards
Institute
1430 Broadway
New York, New York 10018

- IEC Publication 122-1 (1962) Quartz Crystal Units
for Oscillators
 - Section 1, Standard Values and Conditions.
 - Section 2, Test Conditions
 - including Amendments 1, 2 and 3 \$16.20
 - Amendment 1 (1967) \$ 1.40
 - Amendment 2 (1969) \$ 1.00
 - Amendment 3 (1971) \$ 1.80
- IEC Publication 122-2 (1962) Section 3,
Guide in the Use of Quartz Oscillator
Crystals, including Amendment 1 \$15.40
 - Amendment 1 (1969) \$ 1.40
- IEC Publication 122-3 (1962) Section 4,
Standard Outlines, including
Supplements 122-3A, 122-3B and 122-3C \$23.60

IEC Publication 283 (1968) Methods for the
Measurement of Frequency and Equivalent
Resistance of Unwanted Resonances of
Filter Crystal Unit \$ 6.00

IEC Publication 302 (1969) Standard Definitions
and Methods of Measurement for Piezoelectric
Vibrators Operating Over the Frequency Range
up to 30 MHz \$10.80

IEC Publication 314 (1970) Temperature Control
Devices for Quartz Crystal Units, including
Supplement 314A \$25.00

- Contents. General Characteristics
& Standards;
- Test Conditions; Pin Connections

IEC Publication 314A (1971) First Supplement
to Publication 314 (1970) \$12.60

- Contents. Guide to the Use of
Temperature Control Devices for
Quartz Crystal Units.

IEC Publication 368 (1971) Piezoelectric Filters,
including Amendment 1 and Supplement 368A
and 368B \$45.80

- Contents. General Information &
Standard Values; Test Conditions

IEC Publication 368A (1973) Supplement to
Publication 368 (1971) \$17.40

- Contents. Guide to the Use of
Piezoelectric Filters

IEC Publication 368B (1975) Second Supplement to
Publication 368 (1971) \$16.00

- Contents. Piezoelectric Ceramic
Filters

IEC Publication 444 (1973) Basic Method for
the Measurement of Resonance Frequency
and Equivalent Series Resistance of Quartz
Crystal Units by Zero Phase Technique in a
 π - Network \$13.80

IEC Publication 483 (1976) Guide to Dynamic
Measurements of Piezoelectric Ceramics
with High Electromechanical Coupling

Department of Defense

Order through: Naval Publication & Form Center
5801 Tabor Avenue
Philadelphia, PA 19120

MIL-C-3098 Crystal Unit, Quartz, General
Specification For

MIL-H-10056 Holders (Enclosures), Crystal, General
Specification For

MIL-STD-683 Crystal Units, Quartz; And Holders,
Crystal

MIL-C-39020 Crystal Units, Quartz Established
Reliability, General Specification For

MIL-F-28734 Frequency Standards, Cesium Beam,
General Specifications For

MIL-O-55310 Oscillators, Crystal, General
Specification For

MIL-F-18327 Filters, High Pass, Low Pass,
Band Pass Suppression and Dual
Functioning, General Specification For

MIL-O-39021 Oven, Crystal, General
Specification For

622 blank